Vol. 61, No. 6 November, 1954

Psychological Review

THEODORE M. NEWCOMB, Editor
University of Michigan
Lorraine Bouthilet, Managing Editor

CONTENTS

Traumatic Avoidance Learning: The Principles of Anxiety Conservation and Partial Irreversibility. Richard L. Solomon and Lyman C. Wynne 353

The Psychoanalytic Theory of Conflict: Structure and Methodology. Allen T. Dittmann and Harold L. Raush 395

A Decision-Making Theory of Visual Detection. Wilson P. Tanner, Jr. and John A. Swets 401

On Abelson's Critical Comment. David Rapaport 410

PUBLISHED BIMONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

CONSULTING EDITORS

SOLOMON E. ASCH
ROBERT R. BLAKE
STUART W. COOK
CLYDE H. COOMBS
LEON FESTINGER
W. R. GARNER
JAMES J. GIBSON
D. O. HEBB
HARRY HELSON
E. R. HILGARD
CARL I. HOVLAND
E. LOWELL KELLY
DAVID KRECH
ROBERT W. LEEPER
ROBERT B. MACLEOD
DAVID C. MCCLELLAND
GEORGE A. MILLER
GARDNER MURPHY
OSCAR OESER
CARL PFAFFMANN
CARROLL C. PRATT
DAVID SHAKOW
RICHARD L. SOLOMON
ELIOT STELLAR
S. S. STEVENS
ERIC TRIST
EDWARD L. WALKER
ROBERT W. WHITE

The *Psychological Review* is devoted to theoretical articles of significance to any area of psychology. Except for occasional articles solicited by the Editor, manuscripts exceeding twelve printed pages (about 7,500 words) are not accepted. Ordinarily manuscripts which consist primarily of original reports of research should be submitted to other journals.

Because of the large number of manuscripts submitted, there is an inevitable publication lag of several months. Authors may avoid this delay if they are prepared to pay the costs of publishing their own articles; the appearance of articles by other contributors is not thereby delayed.

Tables, footnotes, and references should appear on separate pages; all of these, as well as the text, should be typed double-spaced throughout, in all manuscripts submitted. All manuscripts should be submitted in duplicate. Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail. Manuscripts should be addressed to the Editor, Dr. Theodore M. Newcomb, Doctoral Program in Social Psychology, University of Michigan, Ann Arbor, Michigan.

PUBLISHED BIMONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1333 SIXTEENTH ST. N. W., WASHINGTON 6, D. C.

$6.50 volume $1.50 issue

Entered as second-class matter July 13, 1897, at the post-office at Lancaster, Pa., under Act of Congress of March 3, 1879

Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized Jan. 6, 1950

Send all business communications, including address changes, to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must arrive by the 15th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Copyright 1954 by the American Psychological Association, Inc.

VOL. 61, NO. 6 NOVEMBER, 1954

THE PSYCHOLOGICAL REVIEW

TRAUMATIC AVOIDANCE LEARNING: THE PRINCIPLES OF ANXIETY CONSERVATION AND PARTIAL IRREVERSIBILITY

RICHARD L. SOLOMON AND LYMAN C. WYNNE [1]

Harvard University [2]

The purpose of this paper is to describe two particular ideas, which we shall call "anxiety conservation" and "partial irreversibility," within a general theory of anxiety and avoidance learning. In doing so, more familiar postulates appear to us to generate some new and interesting theorems about behavior, some of which seem to correspond to established facts. We shall not concern ourselves at this time with an exhaustive review of empirical data. Rather, we shall direct our attention to a theoretical argument and shall illustrate specific points with observations drawn from the following fields: avoidance learning, psychotherapy, physiological psychology, and psychosomatic medicine.

Although we shall make no attempt here to validate thoroughly the logical deductions from our theoretical notions, we wish to point out that these conclusions are strongly suggested by empirical data reported in a series of our papers (39, 78, 79, 80, 87) as well as in other recent research reports (13, 14, 43, 58, 59, 75).

Pain-fear. It is a behavioral axiom that there are certain classes of stimuli capable of eliciting massive *pain-fear* reactions (see Miller, 59). These classes of stimuli are usually called unconditioned fear stimuli (see Mowrer, 62). The capacity of such unconditioned stimuli to produce pain-fear reactions is often presumed to be innately given (59), but some doubt about this has been raised by Hebb (32) in the case of a limited number of unconditioned stimuli. Whether innately given or not, the essential components of a massive pain-fear reaction may be characterized for analytic purposes as follows: (*a*) Autonomic nervous system discharge, resulting in visceral responses of high magnitude which are followed by feedback stimulation arising in the viscera and affecting the central nervous system; (*b*) skeletal motor discharge, resulting in diffuse and vigorous skeletal reactions which are followed by proprioceptive feedback stimulation arising in the musculature and joints and affecting the central nervous system; such reactions include aversive movements or

[1] We wish to thank our colleagues for many instructive and constructive criticisms of the preliminary drafts of this manuscript. We are especially indebted to G. W. Allport, Stanley Cobb, D. O. Hebb, W. O. Jenkins, G. L. Lindzey, O. H. Mowrer, and Eliot Stellar for valuable suggestions and criticisms.

[2] Much of the speculation in this paper leans heavily on impressions gathered from a project on anxiety in the Laboratory of Social Relations, Harvard University. This project is supported through the generosity of the Medical Sciences Division of the Rockefeller Foundation.

escape responses; (*c*) neuroendocrine discharge, resulting in the secretion of hormones which is followed by chemical feedback to the central nervous system and other physiological systems; and (*d*) higher central nervous system activity, as a direct consequence of afferent and efferent activity.

These four classes of phenomena will be assumed to be present whenever pain-fear reactions of high intensity are observed. Since our discussion will be limited solely to behavior which arises from very intense pain-fear stimulation (trauma), the reader may infer the presence of the four characteristics above whenever the words *fear* or *anxiety* are used in the following discussion. The parametric studies needed for the precise definition of intense trauma or intense fear are at present lacking. The absence of such a definition is a major weakness of this essay. However, we are temporarily willing to trust the intuitive judgment of the psychologist. We hope that most readers will have private conceptions of the attributes of intense fear as contrasted with weak fear, and we further hope that there will soon be some agreement in defining such a distinction.

Two conditioning processes. We shall maintain, as have several other writers (see Mowrer, 63; Schlosberg, 70; and Skinner, 77), that the facts of classical conditioning reflect a process which is not the same as the process of instrumental learning. Each process has distinctive characteristics (see Hilgard and Marquis, 33).

As applied to the learning of instrumental avoidance responses in the presence of intense pain-fear, the analysis of the two processes, we believe, should be made somewhat along the general lines suggested by Mowrer (62) in a provocative paper. We shall assume that pain-fear reactions become conditioned to previously neutral stimuli by virtue of a process of *Pavlovian,* or *classical* conditioning. The essential relationships of this type of conditioning as applied to fear reactions would be as follows:

> Any previously neutral stimuli which are followed closely in time by the occurrence of an unconditioned pain-fear stimulus, together with its immediate, elicited fear reaction, will eventually become capable of eliciting a fear reaction without the presentation of the unconditioned stimulus. The latter fear reaction is said to have become conditioned to the previously neutral stimulus (which is now the conditioned stimulus for a conditioned fear reaction). Following an earlier suggestion of Mowrer (61), we shall define the conditioned fear reaction as an *anxiety* reaction. The use of a term other than fear is justified here on at least two different counts. First, the conditioned reaction may have different components and different amplitude when compared to the unconditioned fear reaction; and second, the conditioned fear response is "anticipatory" in relation to the temporal sequence of events by which it is established. We assume the acquisition of an anxiety reaction to follow closely the empirical laws of Pavlovian conditioning.

The second process which occurs in the establishment of avoidance learning is that of *instrumental* conditioning. We may think of this type of conditioning as following either the laws of S-R reinforcement theory, of S-R contiguity theory, or of more recent cognitive learning theory. (For purposes of our exposition, we believe that it will make little difference which theoretical bias one might have.) The terms "anxiety reduction" or "fear reduction" may be translated for the purposes of this paper to signify stimulation change or termination. Applied to the avoidance conditioning situation, the relationships of instrumental conditioning may be described as follows:

> If a skeletal response occurs in the presence of an intense pain-fear reaction, and the characteristics of the response are such that, immediately subsequent to its occurrence, the intensity of the fear reaction is decreased, then

the skeletal response will be reinforced (its probability of occurrence will be increased in the presence of the recurring fear reaction). Thus, any conditioned stimulus which has acquired the capacity to elicit an anxiety reaction can subsequently acquire the capacity to elicit some skeletal act. One can assume that the anxiety reaction produces feedback stimulation with drive stimulus properties (see Miller, **58**, and Mowrer, **63**). Therefore, if some skeletal act results in removal of the conditioned stimulus for an anxiety reaction, the skeletal act will be reinforced. Anxiety reduction can be thought of as the reinforcing event for the skeletal act. But, before there can be any anxiety to reduce, the initial fear reaction in the presence of some unconditioned stimulus must have been conditioned to some neutral stimulus. An act which removes such a stimulus from the environment will decrease a conditioned anxiety reaction, and thus will be reinforced as an instrumental avoidance response.

Two sets of experimental conditions. The two types of conditioning require different conditions for the establishment of strong responses. The conditions which should be met for the *establishment of intense anxiety,* governed by the laws of *classical* conditioning, are as follows: (*a*) The intensity of the unconditioned, fear-eliciting stimulus, and the initial pain-fear reaction to it, must be great; (*b*) there must be reasonable temporal contiguity between the occurrence of the conditioned stimulus (previously neutral) and the occurrence of the unconditioned stimulus; and (*c*) the CS-US sequence probably must be repeated several times.

It should be emphasized that the three conditions above are needed for the establishment of high-amplitude (intense) anxiety in the presence of some specific conditioned stimulus. If something less is desired, certain aspects of the conditions above may be eliminated or altered. However, in our present discussion, we are only interested in anxiety reactions of a very intense nature. We are not convinced that repetition of the CS-US sequence is absolutely necessary for the establishment of an intense anxiety reaction in the presence of some conditioned stimulus, provided that the intensity of the original pain-fear reaction to the unconditioned stimulus has been very great (severe trauma). The number of repetitions necessary may also be affected by the "perceptual vividness" of the CS. However, we do know that if the three conditions above are carefully met, we will be able to establish a strong anxiety reaction occurring in the presence of some previously neutral stimulus.

The conditions which must be met for the establishment of very strong instrumental avoidance responses are as follows:

1. A skeletal act can terminate both the conditioned stimulus (which is a "signal" of approaching trauma) and the unconditioned stimulus; and furthermore, this skeletal act is of such a nature that it can *terminate* the conditioned stimulus before the unconditioned stimulus is presented, and it can *prevent* the occurrence of the unconditioned stimulus in the regular CS-US sequence.

2. The skeletal act must be followed closely in time by either *pain-fear* reduction (when the organism is "escaping" the unconditioned stimulus) or *anxiety* reduction (when the organism is "escaping" from the conditioned stimulus and is "avoiding" the unconditioned stimulus).

There is much experimental evidence which does not run counter to the analysis we have given so far. The work of Masserman (**56, 57**), Mowrer (**63**), Miller (**58**), Maier (**55**), and Liddell (**44**) provides many examples of the establishment of conditioned anxiety reactions *accompanied by* the development of escape or avoidance responses. While the exact interdependence of the two types of phenomena during acquisition is not yet known,

Solomon and Wynne (80) have shown that conditioned anxiety reactions are apt to appear prior to the emergence of successful avoidance reactions. This was also observed by Mowrer and Lamoreaux (65). Such seeming independence of the two phenomena often reflects itself in very sudden acquisition of avoidance learning under the impetus of severe trauma. For example, Solomon and Wynne (80, 87), Brush, Brush, and Solomon (15), and Kamin (39) have shown several instances of a sudden transition from escaping to avoiding rather than a gradual shortening of latencies to the CS. Kimble (40) has reported the same phenomenon in rats. An example in Fig. 1 from the data of Solomon and Wynne shows this clearly. This dog quickly learns to escape from the shock, demonstrates a plateau, then suddenly acquires a stable, short-latency avoidance response.

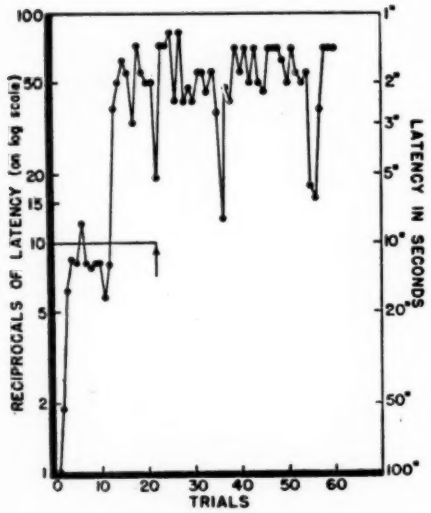

FIG. 1. A typical "sudden learner," from the data of Solomon and Wynne (80). The reciprocals of latency of response to the CS (in seconds) are plotted as a function of trials. The horizontal line marks the CS-US interval of ten seconds; points below this line are escape responses and points above it are avoidance responses. The arrow designates ten avoidances in a row. Note that this dog learned to escape quickly, achieved an escape plateau, then suddenly learned to avoid the shock by responding to the CS in less than ten seconds. Note the stability of the short-latency avoidance responses. They were even more stable after 200 extinction trials.

Avoidance learning is facilitated when the instrumental act not only prevents the occurrence of the US (avoids it) but also terminates the CS or danger signal (65). On the other hand, Bitterman, Reed, and Krauskopf (10) have shown that delay of termination of the US following an instrumental avoidance reaction (delay of reward) tends to strengthen rather than weaken the avoidance response. In their experiment they found that long shock duration is more important in intensifying the classical conditioning process than it is in weakening the growth of the instrumental response by virtue of delay of reinforcement. These authors consider their data to be strongly consonant with a dual-process theory of avoidance learning. Gantt and his collaborators have also noted the relative functional independence of anxiety reactions and instrumental acts (21, 26, 27).

We have no really strong convictions about the adequacy of S-R reinforcement theory, S-S contiguity theory, or S-R contiguity theory in handling the facts of the *development* of anxiety and avoidance responses. It will, we believe, occur to many that these theories may be applicable to many selected aspects of the data. But we do feel that the two-process approach to an analysis of anxiety and avoidance learning has helped us greatly in organizing for our own purposes a large body of experimental evidence. This, we feel, will become more apparent in handling the facts of *extinction* of traumatic avoidance learning.

AN ANALYSIS OF EXTINCTION OF INSTRUMENTAL AVOIDANCE RESPONSES

The strength of the learned instrumental avoidance response will presumably be related to the intensity of the classically conditioned anxiety reaction. Once the instrumental avoidance response is occurring regularly in the presence of the conditioned stimulus, then of course the unconditioned stimulus is omitted, and the temporal contiguity between the CS and the US is destroyed. When this occurs, the organism is terminating the conditioned stimulus instrumentally. It is important to note that once the organism is avoiding the unconditioned stimulus regularly, and so is not receiving any traumatic stimulation, we are meeting the conditions usually believed to be required for the *extinction* of a *classically* conditioned response. That is, the conditioned stimulus is no longer followed by the unconditioned stimulus, and so one would expect according to Pavlovian laws that the conditioned anxiety reaction would gradually extinguish. If extinction of the classically conditioned anxiety response actually occurred, then we would ordinarily expect the appropriate instrumental avoidance response to extinguish sooner or later. These events are more or less expected in terms of a two-process theory; but unfortunately, in traumatic learning, they often do not occur as prescribed.

Failure of extinction is difficult for any theory to handle. However, this phenomenon does occur, especially in traumatic avoidance learning. Sometimes the experimenter is impatient, and so sometimes the subjects in such experiments are merely characterized as having a *high resistance to extinction*. Whether one is willing to wait through the hundreds of trials often required for extinction of avoidance will determine one's views on the extinguishability of the instrumental acts. At any rate, it is clear that in traumatic learning we often do face some special problems in extinction. Mowrer (63) has reported extreme resistance to extinction of avoidance responses in rats; so have Miller (59) and Gantt (27). Masserman (56) and Maier (55) have made some related observations which corroborate Mowrer's impressions.

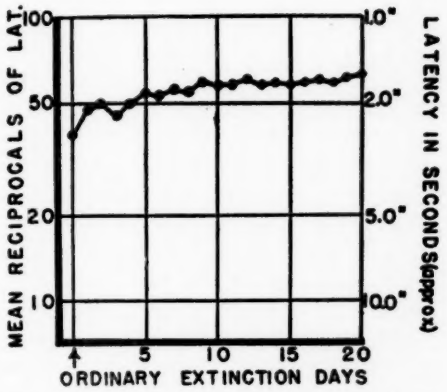

FIG. 2. A mean extinction curve from the data of Solomon, Kamin, and Wynne (78). Response latency as a function of unreinforced (no shock) trials. There were 10 extinction trials on each day.

More recently, Solomon, Kamin, and Wynne (78) have made a rather molecular analysis of the behavior of dogs during extinction procedures following traumatic avoidance learning. They have shown that latencies (reaction times) of avoidance responses continue to shorten after the avoidance response occurs regularly. Thus, with steady omission of the US, the instrumental response of jumping a barrier became more and more stereotyped and the latency of response to the CS became more rapid (shorter), leveling off at 1.6 seconds. Dogs typically continued to respond to the CS for several hundred trials without signs of extinction. Figure 2 shows a mean extinction curve for 13 dogs. Some of these dogs had

received only three or four shocks during a rapid acquisition sequence with a 10-second CS-US interval. The shock was extremely intense, just subtetanizing, during acquisition. Even with relatively long CS-US intervals, Brush, Brush, and Solomon (15) have shown that dogs in traumatic avoidance learning will "settle down" to short-latency responses to the CS in 200 trials after the last shock has been administered during acquisition. For example, with a 20-second CS-US interval during acquisition, the response latency asymptote approached 1.6 seconds, on the average, after 200 extinction trials.

Working with the writers, Kamin (39) has further shown that "spontaneous" jumping, in a free-responding avoidance learning situation, is most intense *after* the criterion of acquisition is reached and short latencies of response to the CS are observed. If frequent spontaneous instrumental acts can serve as a rough anxiety index, it is clear that cessation of presentation of the US is no guarantee of reduction of the strength of the conditioned anxiety reaction. Defecation, urination, and other ANS signs accompanied spontaneous jumping.

From our own data, and from those of others, it therefore seems clear that the extinction data of traumatic avoidance learning cannot be explained solely by existing principles. Existing theories would probably place emphasis on number and regularity of reinforcements, amount of anxiety reduction, lack of reality testing, strong expectations, stereotyped S-R contiguity, etc. But protracted extinction is not indigenous to any popular theoretical system. Two-process theory, as advocated by Mowrer (62), actually does contain some inkling of the resistance phenomenon, but it is not completely adequate in handling failure of extinction or certain observations or overt "emotionality." Gantt's (27) reasoning accepts the phenomenon of resistance to extinction but does not seriously try to explain it.

The expectations of the two-process theory of avoidance learning are portrayed in Fig. 3. Here we see that anxiety reduction reinforcement strengthens the instrumental avoidance response until anxiety is extinguished. Anxiety extinguishes because the CS is no longer followed by the US after the animal is successfully avoiding shock. Then, when the CS can no longer elicit anxiety, the instrumental act extinguishes because it is no longer followed by anxiety reduction. Thus, while Mowrer's notion does protract extinction somewhat, the phenomenon still should be readily obtainable. It may be that the fictitious state of affairs in Fig. 3 is a good model for *nontraumatic avoidance learning*. There are, however, some observations made by Solomon and Wynne

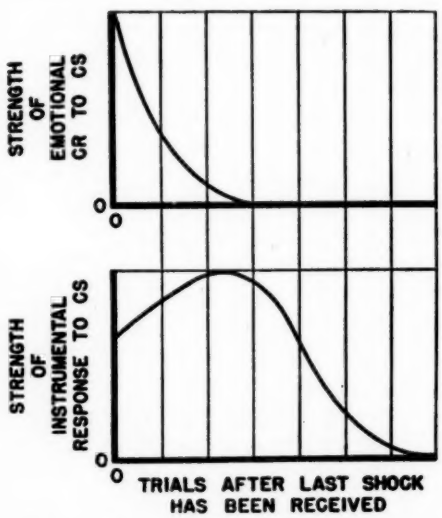

FIG. 3. Hypothetical relationships between intensity of a conditioned anxiety reaction and a learned instrumental response which terminates the CS for the anxiety reaction. These relationships apply to extinction trials only, and so the abscissa is in trials after the last shock reinforcement has been presented.

(80) and by Solomon, Kamin, and Wynne (78) which would render the simple two-process model inadequate for traumatic avoidance learning. These are the data to be explained:

a. We have noted the fact that, in return for a few intense shocks during acquisition of avoidance, dogs gave back as many as 650 avoidances without showing any signs of extinction. Others have observed essentially the same phenomenon (26, 44), and clinical evidence of a similar nature on human phobias exists in great abundance.

b. More important is the fact that overt signs of anxiety rapidly disappeared while the dogs were becoming more and more stereotyped in their jumping and their latencies to the CS were shortening. (If anxiety was being reduced by jumping, the anxiety reduction certainly was not evident at that stage.)

c. If anxiety was occurring covertly in the central nervous system, it did not obey any common laws of extinction, because if the dogs were forcibly prevented from jumping by means of a glass barrier they usually showed intense overt anxiety reactions (78).

d. If a dog happened to have an abnormally long latency on a particular trial, he typically acted "upset" immediately *after* the instrumental response had occurred, and jumped very quickly on the next few trials.

We do not think any popular theory could rigorously deduce these four phenomena; and our current favorite, two-process theory, will *not* do so. We feel that there are at least two important principles which might operate in producing all of these phenomena. We are not certain whether they are supplementary or independent principles. Let us say, then, that they are merely tentative ideas which will make two-process theory work better in handling the four facts of traumatic avoidance learning which we have selected and listed above.

The anxiety conservation phase. There is at least one very important possibility which has been completely ignored in the current analysis of extinction of avoidance responses. After the subject is responding to the CS with latencies shorter than the time required for the elicitation of the classically conditioned anxiety reaction, it is quite possible that at least the peripheral ANS part of this reaction will not occur at all. We hypothesize that the subject removes himself from the presence of the CS so rapidly that the CS is *almost* ineffective. The extent to which this is possible or likely for non-ANS components of anxiety reactions will be discussed later, in the section on physiological problems related to these ideas.

Figure 4 shows how intensity of the anxiety reaction might vary with time of presence of the CS, and how the peripheral anxiety reaction might subside after CS termination, as a function of time of presence of the CS. If the hypothetical events in Fig. 4 are approximately valid, then the occurrence of a rapid instrumental response to the CS would prevent peripheral anxiety reactions from occurring; *if nonreinforced exercise of a CS-CR relationship is the necessary condition for extinction,* then the extinction of the associational linkage between the CS and at least this portion of the anxiety reaction cannot take place. In one sense, the amplitude of the anxiety reaction is being *conserved* as a relatively intact potentiality, a latent functional entity. In common sense terminology, the subject is responding so quickly in the presence of the danger signal that he removes himself from its presence before he can become upset by it. Thus he never overtly experiences the set of events

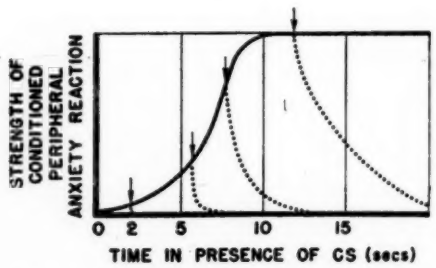

FIG. 4. Hypothetical relationships between time in the presence of an appropriate CS and the postulated strength of the ensuing conditioned emotional reaction. The arrows designate four points at which an instrumental act terminates the CS, and the dotted curves show the form of the anxiety decay curve for those four points. The term "strength" here is both quantitative and qualitative: it is assumed that more and more response elements of the ANS are recruited into the emotional reaction pattern as time in the presence of the CS increases. Thus, as time in the presence of the CS increases, the probability of long-latency visceral reactions being conditioned to the CS will increase.

necessary to "test reality." The anxiety reaction rarely occurs.

Under these conditions, however, the instrumental act *is not followed by anxiety reduction,* and so its habit strength will begin to decrease. When this decrease in habit strength of the instrumental avoidance response reveals itself in longer latencies, then the CS once again has enough time to elicit an anxiety reaction, and the instrumental avoidance response this time will be followed by anxiety reduction and a resultant increment in habit strength. Note how the long latencies in Fig. 1 are followed by rapidly shortening ones. In other words, the subject has "frightened himself" by not getting away from the danger signal fast enough, and he is "relieved" when he finally does remove himself from the signal. Simultaneously, however, there is weakening in the strength of the classically conditioned anxiety reaction, because the US has not occurred following the CS and the appearance of anxiety. The extent to which this weakening occurs will depend in part upon the intensity of the peripheral anxiety reaction, which, as shown in Fig. 4, is possibly a function of the time spent in the presence of the CS (the latency of the instrumental avoidance response).

It is interesting to note that the delay in the presence of the CS constitutes a *partial reality testing experience,* the effectiveness of which is probably proportional to the intensity of the emotional reaction that occurs during the delay. In a real sense we can say that drive strength, D, in Hull's (34) sense, is being "traded" for habit strength, ${}_{S}H_{R}$, as a consequence of such partial reality testing. During short, stable latencies of the instrumental avoidance response, habit strength is being sacrificed while drive (anxiety) is being conserved.

So far, then, we have accounted for facts *b, c,* and *d* above. But troublesome fact *a* still remains. In view of the foregoing theoretical considerations *extinction should occur* sometime, even though painfully resistant to ordinary extinction procedures. At some hypothetical moment, the CS should no longer elicit the anxiety reaction (owing to frequent partial reality testing) and then the extremely strong ${}_{S}H_{R}$ should begin to be overcome by ${}_{S}I_{R}$ (borrowing from Hull). The instrumental response is at this point no longer being reinforced by drive reduction. This is probably contrary to fact *a*.

Therefore, we do not really believe that ordinary extinction procedures must be effective in the case of severe trauma. Suppose, for a moment, we assume that traumatic avoidance learning, if terrifying enough, is completely resistant to *ordinary extinction procedures;* that, barring accidents in procedure, it is empirically possible to produce avoid-

ance responses which will last for thousands of trials over a period of years. Our own observations (78) lead us to believe that this is, in fact, to be expected in dogs, though we are sheepishly aware of the fact that we haven't had the courage to stick with a dog for more than a few months of steady responding. Observations of Maier (55), Masserman (56), Liddell (44), Mowrer (63), and Gantt (26) bolster our feelings about this failure of extinction. So do recurring observations of persistent phobias in animals and man. Therefore, there must be a point at which the anxiety conservation phase is buttressed in some way; there must be some reason for such resistance to extinction as is represented in fact *a* above.

The principle of partial irreversibility of classical conditioning. In general, we feel comfortable with the analysis so far, *but with one extremely important qualification.* In the case of *intense* anxiety, established with the support of an initial, *intense* pain-fear reaction, we believe that the classically conditioned responses *have become incapable of complete extinction.* We are assuming that in such cases of intense anxiety (conditioned fear reactions), where the autonomic and skeletal reactions are of *great magnitude and involvement,* a *principle of partial irreversibility of classical conditioning* is operating. The principle we wish to describe is not that of total irreversibility, but rather that a "traumatic" or very intense "pain-fear" reaction taking place in the presence of some conditioned stimulus pattern will result in a *permanent* increase in the probability of occurrence of an anxiety reaction in the presence of that conditioned stimulus pattern (whenever it reoccurs).

This permanent change can be thought of as a decreased threshold phenomenon or as a sensitization phenomenon which is relatively permanent. A neurophysiological correlate might be the permanent reorganization of central nervous system networks. (These possibilities will be discussed later.) A physical analogy would be hysteresis. Such conceptions of partially irreversible changes (underlying behavioral phenomena) bear a strong relation to the concept of "the adaptation syndrome" which has been described by Selye and his colleagues (73). The adaptation syndrome is characterized by relatively permanent, partial reorganizations of hormonal functioning. We are merely generalizing the principle of partial reorganization from endocrinology to neurophysiology, and then to behavior.

Certainly this *general* notion is not original! Freud (23) uses the idea of partial irreversibility in describing the effects of trauma. (In fact, Freud also has an anxiety-preserving mechanism, repression, to go along with partial irreversibility. But perhaps this analogy is stretched.) Hull's (34) concept ${}_{S}H_{R}$ is established through the action of reinforcement (drive reduction), and it is preserved *as a fixed quantity.* Extinction is predicted by the fact that ${}_{S}I_{R}$ and I_{R} are subtracted from ${}_{S}H_{R}$ in the computation of ${}_{S}E_{R}$, effective reaction potential. White (86) and Allport (3) certainly would be astonished if a general concept of irreversibility were held to be novel. So, perhaps, might be Maier (55), Hebb (32), Kubie (41), and a host of others. But we would like to apply the partial irreversibility principle in a *very particular manner,* reserving it as a *property of the classical conditioning of "intense anxiety" reaction.* We feel that this specification of the general principle is in accord with evidence from a variety of sources and may be useful in generating psychological hypotheses which are more directly testable than has been possible with the *general* idea of partial irreversibility.

If this principle of partial irreversibility is taken seriously, it would mean that the *ordinary extinction procedures* of Pavlovian conditioning, when applied to a conditioned anxiety reaction of great intensity, will have *only a limited effectiveness in reducing the intensity of such a conditioned response.* The extinction procedures (characterized by dissociation of the CS and US) of classical conditioning should only be capable of diminishing the strength of a conditioned anxiety reaction down to some *irreducible minimum* in the presence of the CS, to some fixed threshold value which is above the zero point. The anxiety reaction will take place to some extent in the presence of a protracted conditioned stimulus; and even though extinction procedures are repeatedly employed over long periods of time, the anxiety reaction will merely decrease somewhat in intensity, never completely disappearing. Thus, the ordinary extinction procedures of *classical* conditioning will be only partially successful (see Fig. 5). The conditioned stimulus will always have the capacity to elicit an anxiety reaction of some magnitude. The irreducible, minimum, elicitation capacity of the conditioned stimulus will probably be a function of the intensity of the conditioned anxiety response before extinction procedures are started.

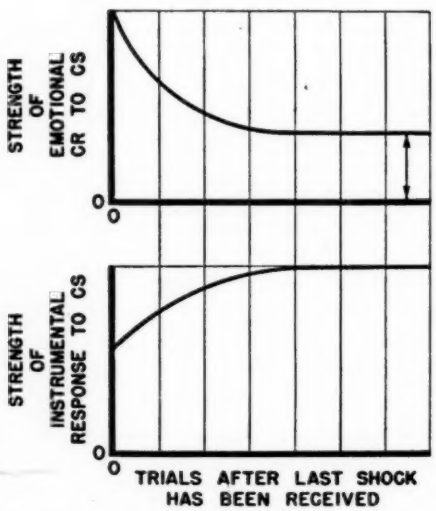

FIG. 5. Hypothetical relationships between the intensity of a conditioned anxiety reaction and a learned instrumental response for *traumatic* avoidance learning. Note that the asymptote for the emotional reaction extinction curve is above zero. The arrow designates the irreversible increment. Note that the strength of the instrumental avoidance response approaches some asymptote at a high value.

When we consider the significance of these assumptions for a theory of avoidance conditioning, a striking implication emerges. Without tampering with either the law of effect, the principle of anxiety reduction, or S-R contiguity principles, we can predict the *failure of extinction of instrumental avoidance* responses which have been established in the presence of intense pain-fear or anxiety. We arrive at such a conclusion, not on the basis of a drastic reformulation of instrumental learning and extinction laws, but rather, on the basis of partial irreversibility of *classically* conditioned anxiety reactions. There is no shortcoming of the law of effect implied here! If the latency of the instrumental avoidance response is long enough, the conditioned stimulus will elicit some degree of anxiety. Since the instrumental avoidance response has been established in the presence of anxiety, and since the instrumental avoidance response will continue to be followed by reduction of anxiety (by removal of the conditioned stimulus), the instrumental avoidance response will not be weakened. In fact, it will be strengthened through the action of the law of effect, approaching an asymptote of response strength long after extinction procedures have been instituted, long after the organism is successfully avoiding the unconditioned stimulus. Such a case is shown in Fig. 5. The prin-

ciple of anxiety conservation leads us to expect that overt signs of anxiety will be expected to diminish, because the animal is not staying in the presence of the CS long enough to allow them to be elicited. This analysis now fits most of the troublesome facts which we have encountered in our own research (see facts *a*, *b*, *c*, and *d* above, as well as Fig. 2).

RETRAINING TECHNIQUES

Reality testing. Our theoretical fantasies also imply that forced reality testing extinction procedures may fail to eliminate the instrumental avoidance acts. Forced reality testing procedures are characterized by special manipulation, such that the organism is detained in the presence of the conditioned stimulus for long time intervals without the unconditioned stimulus being presented; and in addition, the organism is, by special means, prevented from exercising the learned *instrumental* avoidance response. It has been argued that this "forcing" procedure "lets the organism know" that the conditioned stimulus no longer "signalizes" the future occurrence of the unconditioned stimulus. Forced reality testing should result in "reorganization of the cognitive field" such that the organism no longer "perceives the situation as dangerous." Therefore, it has been argued, such a procedure should be effective in eliminating anxiety in the presence of the conditioned stimulus, and therefore, the instrumental act should subsequently weaken. However, application of the principle of partial irreversibility to forced reality testing leads us to predict that, while the anxiety reaction will be reduced in intensity, it will never be completely removed, and the avoidance response may therefore reappear even after long periods of reality testing.

We wish to emphasize that this theoretical position is *not* a doctrine of therapeutic hopelessness. We have pointed out that the originally intense anxiety reaction, elicitable by a conditioned stimulus, can be substantially reduced by the employment of traditional extinction procedures as well as the therapeutic procedures of reality testing. Some amelioration of anxiety is assumed to be possible, even in the case of most severe trauma. But a particular instrumental avoidance reaction, originally stemming from a severe traumatic event, will tend to persist if we merely employ ordinary extinction procedures or therapeutic reality testing.

Reward or "support." Therefore, it seems clear that if we are to eliminate a strong, learned *avoidance response* we must introduce some new stimulus conditions along with the other procedures above. But in doing so, we must keep in mind that the organism will, despite extinction procedures, always exhibit an anxiety reaction of some sort in the presence of the conditioned stimulus. One requirement would be that any added stimulation should have a high probability of eliciting a pattern of *skeletal responses*, which is *incompatible* with the occurrence of the avoidance response. Fatigue, competing avoidances, and competing appetitive reactions could be utilized. The stimulus pattern could be "motivational" in nature or could be composed of conditioned stimuli controlling strong responses which are incompatible with the avoidance reaction we wish to eliminate. Once the new stimulation conditions are introduced, and the resulting incompatible behavior is reinforced, the probability of the occurrence of the old avoidance response will decrease. The organism will then be capable of performing a new instrumental response in the presence of the conditioned stimulus for anxiety, yet the anxiety reaction *will continue to*

occur. The old phobic or compulsive response will be gone, but not the anxiety reactions. At least, that is what we would expect if the principle of partial irreversibility of classical conditioning were valid.

Use of reward learning, or of appetitive motivation, in the elimination of strong avoidance responses has been studied to some extent, but the results are confusing. On the one hand, we have the excellent studies of Lichtenstein (43), in which traumatic stimulation was used to produce lasting feeding inhibition in dogs. And, on the other hand, we have those famous accounts of the curing of a child of fear of furry animals by bringing the feared object into the feeding situation (37, 38). Our guess is that the latter work represented a certain combination of forced reality testing, "crowding the threshold" of the anxiety reaction, *and* good luck! Unless the actual degree of anxiety is known, and the relation of its strength to that of hunger is known, it would be possible to produce neurotic inhibition of eating in the child by introducing the conditioned stimulus. This problem needs to be explored, especially with respect to avoidance responses that are established on the basis of intense trauma and overwhelming fear.

Our own observations are equivocal here. We have tried to retrain dogs in our traumatic learning situation by keeping them hungry five days and trying to use a feed lure to compete with the instrumental avoidance response. This was not a very successful procedure. Often the dog grabbed at the food and then performed the avoidance response. Much experimentation is needed in this area; especially needed are parametric studies which pit drives against each other in varying strengths, studies which might yield "isomotive curves."

Punishment. Another type of "new stimulation" which might be introduced in order to discourage the occurrence of an avoidance response is ***punishment***. That is, when the organism responds to the conditioned stimulus with an avoidance response, some traumatic stimulus could be applied, preferably immediately following the occurrence of the avoidance response. Our theoretical ideas would deduce that such a procedure would only be effective *if* it is preceded by, or accompanied by, extensive forced reality testing. If punishment *for* avoiding is used before the classically conditioned anxiety reaction is partially extinguished, then the avoidance response might be substantially ***strengthened*** rather than weakened. Punishment would raise the general anxiety level, and since the avoidance reactions have consistently taken place in the presence of anxiety, it is conceivable that the organism would appear to be reacting more strongly to the conditioned stimulus than was the case before the introduction of punishment. However, it is conceivable that punishment could be introduced *after* forced reality testing, or after the anxiety reaction has been reduced to a low level, and that the action of punishment in this case might eventually lead to the extinction of the avoidance response. In this case, we would expect an initial strengthening of the avoidance response followed by gradual elimination of such instrumental responses. We are willing to believe that punishment which directly follows instrumental acts will tend to weaken the instrumental response strength (habit strength) of those acts. But in the case of punishment *without* forced reality testing, the avoidance response is being ***negatively reinforced*** on the one hand and ***positively elicited*** by anxiety drive on the other hand, and so it may appear to be unchanged or, possibly, strengthened by the action of punishment. In the case

TABLE 1

MEAN RECIPROCALS OF RESPONSE LATENCY FOR THE FIVE RESPONSES PRECEDING AND FOLLOWING THE INTRODUCTION OF PUNISHMENT FOR AVOIDANCES

(From Solomon, Kamin, and Wynne [78])

Reciprocals of Latency	Punishment Introduced	
	After 200 Avoidances	After 20 Avoidances
Before punishment is introduced	50.7	50.5
After punishment is introduced	72.9	63.1

of punishment *with* forced reality testing, the anxiety level may be low enough so that the weakening effects of negative *reinforcement* (the Thorndikian stamping-out effect) can emerge after an initial period of raised anxiety level, resulting eventually in the extinction of the avoidance response.

Gwinn's experiment (30) demonstrates the expected phenomena quite clearly. In an escape learning experiment he found that, when shock was no longer given, a group of rats gradually ceased running. But an experimental group, shocked *for* running during extinction, at first ran *faster* for several trials. Some of these animals extinguished eventually, others did not. More recently, Solomon, Kamin, and Wynne (78) have shown the same phenomenon with dogs. In an avoidance conditioning extinction series, dogs were shocked *for* performing the instrumental avoidance response. They became extremely "upset," and performed the avoidance response with significantly *shorter* latencies. These data are shown in Table 1 for trials before and after the introduction of punishment for responding. In the same paper, it was reported that more satisfactory retraining occurred when punishment followed reality testing procedures than when punishment came before reality testing procedures.

RELATED PROBLEMS IN THE PSYCHOLOGY OF LEARNING

Escape. One special case of learning initiated by noxious stimulation has some characteristics differing from those of avoidance learning. This learning type has usually been called escape learning or conditioning (33). It is characterized by simultaneous presentation of the conditioned and unconditioned stimuli, so that any instrumental acts on the part of the organism can never result in pain avoidance, merely in pain termination. Sheffield (74) has made a keen analysis of the events of escape learning contrasted with those of avoidance learning, and we are in essential agreement with his conclusions. He shows that in escape learning the presentation of the unconditioned stimulus tends to elicit unconditioned responses that are incompatible with the performance of a specific instrumental escape response which happens to terminate the CS and the US. For example, if a rat is required to run in order to terminate shock, and the shock level is fairly high, unconditioned crouching reactions may interfere with running. In addition, the rat may terminate shock only if he runs a given distance, so that the instrumental acts such as the initiation of running movements are actually punished for a finite period of time. Therefore, one would expect the development of a lot of diffuse responses (see Schlosberg, 70), some of which would actually be incompatible with a discrete and efficient instrumental escape act. When the unconditioned stimulus is later omitted in the extinction of the escape response, there would be present many response alternatives which might interfere with the previously reinforced instrumental act. Thus, according to Sheffield, the

escape response would be weak, and might be easily extinguished when the unconditioned stimulus is omitted.

However, we would like to add that the general and diffuse reactions which have been elicited in the presence of anxiety would persist if the original stimulation were traumatic enough. We would expect the conditioned stimulus to continue to elicit some anxiety, just as was the case in avoidance learning, and some types of "aversive" skeletal reactions in the presence of this anxiety might be expected. Even though a very specific instrumental escape response may have been eliminated, the classically conditioned emotional reactions would tend to persist, though with an intensity considerably less than was the case before extinction procedures (omission of the US). Thus, the end picture of escape and avoidance extinction might be very much the same: partial persistence of anxiety, but accompanied by new skeletal reactions. At least, such would be anticipated by the principle of partial irreversibility of classical conditioning. We do not know of an experiment directly designed to demonstrate these phenomena, but such an experiment would be, in principle, a simple one.

Partial reinforcement. Sheffield and Temmer (**75**) have clearly demonstrated that escape learning, with shock level held constant, is less resistant to ordinary extinction procedures than is avoidance learning. They argue that the difference is due to the effects of partial or aperiodic reinforcement. Many animals show an irregular sequence of escape and avoidance trials when they are in the course of acquisition of instrumental avoidance responses. An animal may receive a shock on trial six because it did not respond with a latency less than the CS-US interval, and on the seventh trial it may avoid the shock, responding to the CS alone. Then, on the eighth trial, the animal may receive shock again for responding too slowly to the CS. Such alternation from shock to nonshock trials is common in many experiments. This is tantamount to aperiodic reinforcement if the US is considered to be the reinforcer of the instrumental act.

Now, based on a generalization decrement theory (**36**, **76**) or a response-grouping theory (**64**), it would be expected that an aperiodic reinforcement schedule would lead to high resistance to ordinary extinction procedure when compared with the consequences of regular reinforcement. Escape learning is characterized by regular reinforcement, the US being administered on every acquisition trial. Regular extinction procedure merely omits the US and presents the CS alone. This is an abrupt transition from 100 per cent reinforcement to 0 per cent reinforcement, a condition which Sheffield and Temmer believe is conducive to easy extinction because of a generalization decrement—the animals can easily discriminate a change in the situation. But in avoidance training procedures the change from acquisition to extinction procedure is not as discriminable, and so the animals respond for more trials during extinction. (Another way of saying this is that the avoidance training procedure does not give the animals good cues for reality testing.) Sheffield and Temmer point out that early in ordinary extinction procedure the avoidance responses are actually less vigorous (of lower amplitude) than are the escape responses. These data contrasting escape and avoidance learning are paralleled by data in reward learning. In general, experimental work on partial or aperiodic reinforcement has shown that the irregular sequence of reinforcements leads to slower acquisition and slower extinction. With low

ratios, extinction may well be a long and drawn-out procedure (36, 76).

In view of these considerations, it might be asked why we forsake a simple argument, such as the generalization decrement theory of extinction (76), in accounting for high resistance to extinction in traumatic avoidance learning. There are at least two reasons for our decision:

a. With an intense, traumatic US, the transition from escape responses to avoidance responses may often occur in a sudden fashion. An animal may absorb a few high-level shocks and then begin to avoid with perfection. This doesn't always mean that latency changes are abrupt, even though the escape-avoidance transition is. A good example of this is shown in Fig. 6, taken from the acquisition data of Solomon and Wynne (80). The same phenomenon was observed, even with very long CS-US intervals, by Brush, Brush, and Solomon (15). This is tantamount to a sudden shift from 100 per cent shock reinforcement to 0 per cent shock reinforcement, a condition which typifies the onset of ordinary extinction of *escape* responses. Such animals, learning to avoid very suddenly, do *not* appear to be less resistant to extinction than animals having a more aperiodic sequence of shocks and avoidances during learning. Thus, one cannot say that traumatic avoidance learning is typified by aperiodicity or partial reinforcement to the exclusion of perfect periodicity.

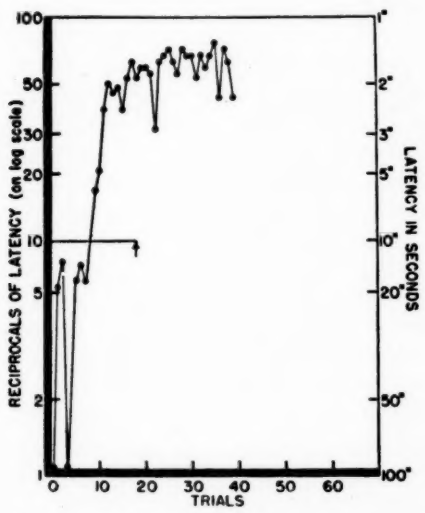

FIG. 6. A typical example of a completely sudden transition from escaping to avoiding, seen in traumatic avoidance learning in dogs (80). Note that while the transition from escape to avoidance is sudden, the latencies in this case change gradually after a sudden shift from a 15-second latency of response to the CS to a 6-second latency. This illustrates that the mere classification of instrumental acts into escape vs. avoidance, or percentage of successful responses, is not a totally adequate representation of the events taking place.

b. Escape learning is not comparable with avoidance learning because the CS-US intervals are not comparable. Escape learning is characterized by overlap of CS and US in time such that the US is unavoidable. But avoidance training procedure places a considerable period of time between CS and US so that the subject can react quickly enough to avoid the US. If the CS-US interval is an important variable in determining resistance to extinction, it would become a more complex task to assess the contribution of pattern of reinforcement as an important variable in traumatic avoidance learning. (This is a problem which we have just begun to explore, and the data are not conclusive at the present time. However, working with the authors, Kamin [39] has shown resistance to extinction to be a function of CS-US interval.)

In view of the data on resistance to special extinction procedures described by Solomon, Kamin, and Wynne (78), Masserman (56), Gantt (27), and others (44, 55), it seems reasonable that in traumatic learning the pattern or sequence of reinforcements is only one contributing variable of many in producing high resistance to extinction.

Nontraumatic learning. It has been suggested that perhaps all learning contains the element of partial irreversibility. Indeed, Allport's concept of functional autonomy (3) may be taken as a prototype for such a position. At present, we do not wish to extend our analysis beyond very intense, traumatic classical conditioning; but we realize that there is a possibility that other types of learning share some of the features of partial irreversibility of classical conditioning. "Law of effect learning," we feel, is totally reversible, and is even extinguishable "below zero" when the correct conditions are met.

In general, we feel that Gantt's (27) argument about the distinctive differences between conditioned autonomic reactions of an "appetitive" nature and conditioned autonomic reactions of the "emergency system" is cogent here. It can scarcely be argued that a Pavlovian conditioned salivary reaction—no matter how associated it might be with intense and eager orientation of the subject, and even overwhelming "expectations" of food (see Zener, 89)—involves the magnitude of visceral and hormonal reorganization that intense conditioned pain-fear reaction does. Nor is it too difficult to extinguish a conditioned salivary reaction below zero, as Pavlov (68) has shown; yet the conditioned heart-rate reaction to intense trauma has been shown by Gantt (27) to be incapable of total extinction in dogs, even over a period of years. While the conditioned appetitive reactions of reward learning may be a strong factor in influencing the resistance to extinction of rewarded instrumental acts, we do not feel that a partial irreversibility phenomenon need be involved.

There is, we feel (yet we cannot at present convincingly prove) a tremendous qualitative difference and quantitative difference between appetitive and traumatic learning. The exact nature of the differences remains to be explored. Perhaps it will point up the problem if we are reminded that a rat may starve to death rather than risk punishment, that a dog will refuse to eat if near a "danger signal" (43), that a sheep may exhibit hypertension due to trauma for months. (Perhaps Cannon [16] was making a distinction eventually useful to the psychology of learning when he delineated the *vegetative* and *emergency* functions of the autonomic nervous system. While this distinction now seems more dubious than Cannon made it out to be, since non-ANS functions operate, it seems premature to forget the distinction.)

RELATED PHYSIOLOGICAL PROBLEMS

During our discussions of the principles of anxiety conservation and of the partial irreversibility of intense anxiety reactions within the framework of learning theory, we have often found it illuminating to consider the compatibility of our ideas with present-day physiological facts and theory. We have felt that the consideration of the anxiety and avoidance learning problem from more than one starting point would suggest surplus meanings, and thus the possibility of additional modes of verification. Three main physiological problems are particularly relevant to this discussion. (*a*) What sort of physiological processes, among those aroused in intense anxiety reactions, are most capable of mediating skeletal avoidance responses? (*b*) To what extent do these physiological processes have discrete latencies, within which a prompt instrumental response could occur or begin? (These two questions pertain directly to the principle of anxiety conservation, which presumes that certain anxiety reactions could, on the one hand, motivate *delayed* avoidance responses, but, on the other hand, could be conserved from extinction by

more *prompt* avoidance responses.) (*c*) To what extent do mediating physiological processes and associated anatomical structures have characteristics making them possibly susceptible to partially irreversible changes in the presence of traumatic stimulation?

As we indicated at the outset of this paper, we like to imagine that during intense anxiety reactions feedback stimulation occurs in at least four major physiological systems: autonomic nervous system, skeletal motor-proprioceptive system, neuro-endocrine system, and higher levels of the central nervous system. Since in the intact organism these systems clearly overlap and interact, and yet serve partly parallel functions, such a classification is highly arbitrary, but may serve expository and heuristic purposes. We shall consider, in turn, each of these four systems with respect to the three questions raised.

Autonomic nervous system and associated feedback. Discharge of the autonomic nervous system has long been regarded as prominent in anxiety. The existence of *efferent* pathways from various CNS levels to the peripheral ANS is now well-known and thoroughly established (24, 47, 84). However, if the autonomic aspects of anxiety are to influence avoidance behavior, anatomical and physiological mechanisms must exist whereby autonomic discharge arouses *afferent* feedback that actually can impinge upon projection areas and eventually influence skeletal motor centers and pathways. Only very recently has there been *direct*, detailed evidence that visceral *afferent* feedback may ascend in the CNS above the level of the lower brain stem (1, 4, 5, 9). It now appears that the neurophysiological characteristics of the visceral afferent projection system "differ in no important respects from the somatic projection other than in peripheral origin" (5, p. 457).

Neural feedback from autonomic reactions probably participates in both the specific and the diffuse projection systems to the thalamus and cerebral cortex. For the *specific* projection system, evidence is accumulating that there is an interaction of visceral and somatic afferent representation at several levels of the CNS (6, 24). Thus, the notion of the two-process learning theory that at least some of the afferent feedback impulses from the viscera have the properties of stimuli and so are capable of becoming conditioned stimuli and drive stimuli now seems reasonably consistent with physiological evidence. That is, not only may visceral reactions become elicitable by external stimuli through classical conditioning, but also the visceral feedback stimuli may become cues for instrumental acts, or may become motivating stimulation if their intensity is great enough (see Miller, 59).

Much recent research indicates that the *diffuse* projection system, involving the reticular activating system of the brain stem, is extremely important in emotional arousal, alertness, and the attentive processes (47, 54). The reticular formation probably also has a descending facilitating influence upon lower motor outflows (48). Although the relative significance of the various kinds of afferent stimulation that can activate the reticular system under different conditions is still not established, it is now very likely that feedback from both sympathetic and parasympathetic systems contributes to this afferent collateral network (22, 88). From the standpoint of avoidance learning theory, it is interesting that afferent impulses to the reticular system do ***not*** seem to function as cues or as specific conditioned stimuli. Rather, they seem to alter diffusely the alertness or attentiveness of the organism in perceiving

and responding to impulses of the more specific projection systems (22, 81).

These physiological considerations, described, to be sure, in a highly schematic and oversimplified fashion, suggest the need for caution by learning theorists in discussing the properties of drive states and especially of acquired drives. Even the relatively well-understood peripheral autonomic nervous system, when it is active, results in afferent feedback impulses having a great potential range of variation in quality and magnitude, in duration and latency, as well as in the extent to which such feedback actually may influence instrumental motor responses. The loose term, "response-produced stimuli," while perhaps useful in highly generalized theory, tends to disguise these issues. And behind such issues may lie processes which greatly affect the extent to which anxiety and other learnable drives may actually mediate instrumental avoidance learning.

This discussion of autonomic feedback has emphasized neural pathways. However, it is clear that ANS discharge also produces widespread hormonal feedback affecting the reactivity of the organism in many diffuse ways. Some features of the possible role of such changes for anxiety and avoidance learning will be pointed out shortly.

To date, the various physiological and anatomical studies that have delineated the *details* of the mechanisms whereby autonomic reactions and feedback could influence instrumental responses have not been conducted under conditions of avoidance learning. Therefore, we do not know the extent to which the possible mechanisms are actually invoked in traumatic avoidance learning as discussed in this paper.

Inferential evidence concerning the *over-all* role of the peripheral ANS under actual traumatic avoidance learning conditions has been presented by Wynne and Solomon (87). They used surgical and pharmacological procedures in dogs to eliminate the sympathetic and parasympathetic peripheral nervous function. This was done in such a way that the capacity to perceive a traumatic US via *direct* sensory paths was unimpaired, and the motor capacity of the dogs to perform the instrumental avoidance response was not affected. In this experiment a CS-US interval of 10 seconds was used (allowing time for ANS reactions to develop in normal controls). Thirteen dogs who were given the surgical-drug procedures before training showed less uniform behavior than did normal control dogs. Ten of the 13 dogs were outside of the range of any normal controls, either in characteristics of acquisition or of extinction. In general, the experimental dogs were often retarded in reaching the avoidance learning criterion as well as in making their first avoidance response. However, they were all capable of achieving the criterion of avoidance learning and of responding with short latencies to the CS. Eight of these 13 dogs extinguished spontaneously. This had never occurred in any of 13 control dogs which had been trained under the same conditions. The experimental animals, as compared with normal controls, showed relative "indifference" to the shock, with few signs of autonomic and motor "upset," and relatively little stereotyping of responses during extinction.

Using tetraethylammonium with rats, Auld (8) has found that the acquisition rate for avoidance learning was depressed and that the extinction rate was accelerated. However, the pharmacological side effects of the drug make it difficult to interpret to what extent these learning changes were initiated by the effects of the drug on the peripheral ANS. A recent experiment by Brady (11) demonstrates these side effects to

be extremely important in depressing general activity.

It was possible for Solomon and Wynne (80) to obtain only very incomplete data on ANS discharge and no objective data on associated feedback in their study of traumatic avoidance learning using normal dogs. Nevertheless, they observed a definite tendency for the first overt ANS reaction to the CS to occur approximately two trials before the first avoidance response. Mowrer and Lamoreaux (65) have reported similar observations. Fully satisfactory experimental evidence on the role of the ANS in contributing to anxiety as a mediating drive state in avoidance learning will not be available until direct measurements of both ANS discharge and associated feedback are possible under avoidance learning conditions varied along standardized parameters. However, the various kinds of inferential data now available quite clearly suggest that the peripheral ANS circuits have at least a highly significant, although not essential, role in avoidance learning under conditions of intense trauma during acquisition.

These various kinds of evidence suggest that, under certain conditions at least, feedback circuits which include the peripheral ANS *facilitate avoidance learning*.

According to the principle of anxiety conservation, the maintenance of avoidance under "ordinary" extinction conditions should depend, in part, upon the relation of the latency of the instrumental avoidance response to the total latency of classically conditioned anxiety reactions (efferent) plus feedback (afferent) which may be capable of facilitating the instrumental response. If we knew the latencies of the relevant physiological processes, we could predict anxiety conservation and prolonged maintenance of avoidance responses at a latency level just below, for most trials, the latency of these physiological processes.

In the case of the autonomic nervous system, the latency of measurable efferent reactions after the presentation of an external CS is ordinarily 1–4 seconds (2, 24, 47). It has been noted, for example, that in association with both light and sound stimuli a galvanic skin response of moderate electrical potential change has a latency of about 1.5 seconds (47). With the same mild stimuli, heart rate changes measured with the electrocardiogram also begin after about 1.5 seconds, reach a maximum after about 3 seconds, and have disappeared after about 4 seconds. Such measurable effector reactions presumably occur only after summation of less overt processes at both central and peripheral levels. These processes include, as examples, local intrinsic nervous reflexes in the gut in conjunction with local chemical and hormonal adjustments, mechanisms such as the carotid sinus reflex affecting cardiovascular and respiratory activity, and complex central circuits involving the primitive forebrain and the reticular activating system of the brain stem.

When the physiological processes brought into activity by external stimuli, especially stimuli arousing intense reactions, are considered in their full complexity, it is obvious that great caution is necessary in interpreting generalized statements about the latency of either efferent or afferent circuits. What can be said with some degree of assurance in the case of the ANS is that, in view of the demonstrable discrete latencies of peripheral ANS reactions, *effective* facilitation of skeletal responses by the feedback returning to central levels must also have a discrete latency. ***Probably at least two seconds, perhaps several seconds, must elapse following a CS before feedback from the peripheral***

ANS can appreciably affect central motor processes.

These physiological latency characteristics of the ANS make the peripheral autonomic portions of anxiety reactions especially likely to be "conserved" by prompt avoidance responses which quickly remove the organism from the presence of the arousing CS. However, conditioned autonomic reactions do occur with a latency of only a few seconds and hence can be expected to help maintain avoidance when the instrumental response is delayed beyond this extent. The physiological properties of the ANS make it necessary to qualify such a generalization in several respects. For one thing, it is not correct to assume that the different parts of autonomic discharge and feedback all have the same latency. This point probably applies more to the latency with which autonomic reactions reach maximum intensity rather than to the latency with which they begin. For example, gastric reactions involving changes in acid production and motility probably take considerably longer to summate to maximum level than do heart-rate changes.

In other words, even within the autonomic group of anxiety reactions, a given duration of exposure to a CS may tend to elicit some kinds of reactions and to conserve others from full arousal. A recent experiment by Kamin (39) sheds a great deal of light on this problem. He studied the role of CS-US interval in a traumatic avoidance learning situation. The emotional reactions of his dogs varied with the length of the interval. Kamin's 5-second group showed a predominance of muscular tension, alertness, defecation, and urination during acquisition of a jumping response. His 20-second group, in contrast, showed a predominance of diffuse, agitated locomotion, retching, vomiting, and stomach rumblings which Kamin happily called "the gastronomic microphonic."

Such a contrast in symptoms must somehow be related to the "natural latencies" of visceral and skeletal emotional reactions. Some of the long-latency reactions probably are never conditioned when the CS-US interval is short. But of those reactions which are effectively conditioned, those with the longer latencies will be conserved during the earlier phases of extinction. It is clear that a detailed study of the qualitative changes in conditioned emotional reactions, during all phases of acquisition and extinction of avoidance learning, is sorely needed.

If partially irreversible changes do occur as a result of traumatic stimulation, such changes are most apt to occur in those structures in which the stimulation is convergent and most intense, or in structures with a particular vulnerability to such changes. Convergence might be especially likely to occur in pathways or areas with many afferent connections but with a relatively primitive or undifferentiated structure. In the case of the peripheral ANS, present anatomical knowledge does not suggest that such convergence is probable.

One might wonder whether the structural changes occurring in psychosomatic illnesses involving autonomic effector organs fall within the scope of the principle of partial irreversibility of intense classically conditioned anxiety reactions. There seems to be general clinical agreement that in individuals with such illness as duodenal ulcer, despite extensive changes in the person's life situation and despite prolonged psychotherapy, the ease of reactivation of the ulcer is greater than in the average individual. The extent to which such vulnerability can be due to traumatic life experiences aside from "constitutional" predisposition is highly con-

troversial and largely speculative at present; "organ inferiority" is a disputed concept in psychosomatic medicine today. Also, it is conceivable that such psychosomatic diseases in autonomic organs might be secondary manifestations of partially irreversible changes in endocrine or CNS structures, rather than primarily in the peripheral ANS as such. This possibility is in accord with preliminary observations of Wynne and Solomon (87) which indicate that elimination of peripheral ANS functions *after* traumatic avoidance learning has already occurred has no effect upon the course of extinction. Dogs in this group were like normal controls in their extremely strong resistance to extinction of avoidance, in contrast to dogs in which the peripheral ANS was eliminated *before* acquisition of avoidance.

Endocrine reactions. Abundant clinical and experimental evidence indicates that endocrine reactions are capable of being classically conditioned. The relationship between emotional stimuli and the secretion of circulating epinephrine by the adrenal medulla has long been known. More recently, it has become apparent that the secretion of the antidiuretic, gonadotrophic, and adrenocorticotrophic hormones from the pituitary gland, as well as insulin from the pancreas, may also be influenced by conditioned emotional states (28). The adrenal medulla and posterior pituitary are controlled by a rich, direct secretomotor innervation (24), whereas the anterior pituitary (and thus indirectly the thyroid, adrenal cortex, and gonads) is probably under the neurohumoral control of the hypothalamus (31, 35). Various experiments indicate that all kinds of stimuli do not necessarily operate by the *same* mechanism to release the pituitary hormones, but that some, such as noise and restraint, require the normal attachment of the pituitary to the CNS, while others may act by way of circulating adrenaline or histamine, or even perhaps, via other metabolites or hormones (28). Thus, some specific understanding is gradually being built up of the various endocrine reactions which might be regarded as part of the pattern of anxiety reactions which accompany traumatic avoidance learning.

In general, the latencies of feedback from endocrine discharge, possibly except for epinephrine, are *considerably longer* than for the neural feedback from the peripheral ANS. Many of the *measurable* changes associated with adrenocortical discharge take one to four hours to reach maximum intensity (28), although the change may *begin* quite promptly. This means that a mediating effect of endocrine feedback within a given learning trial is likely *only if* the CS-US interval is much longer than that used in all reported avoidance learning experiments. However, such long intervals do occur in certain life situations of human beings faced with anxiety-evoking signals. Effective instrumental responses may not be possible for hours, days, or longer. In experimental avoidance learning an "acclimation" period may allow for endocrine reactions to occur to the total experimental situation. Also, with prolonged experimental sessions endocrine reactions aroused by early presentations of the CS may produce endocrine reactions affecting later responses. (May this contribute to so-called warm-up effects?) From the standpoint of avoidance learning theory, such considerations raise the question of whether endocrine feedback should be regarded as having "stimulus" qualities or as altering the reactivity of the organism to other, more specific stimuli.

Exactly how such endocrine discharge may "afferently" modify behavior and learning has not been studied in detail thus far. That a significant effect can

occur is suggested by the frequency with which psychological aberrations are observed in patients with disorders of the endocrine glands (19, 83), and the frequency with which the administration of hormonal preparations results in deviant behavior (17, 69). However, the *specific* role of any hormone in avoidance learning processes is almost completely unknown. The difficulty in obtaining information on specific effects arises in great part from the impossibility of distinguishing where one hormonal regulatory mechanism stops and another starts. Regardless of the resultant behavioral effect, hormones act only by accelerating or retarding the rates of intracellular reactions which are catalyzed by specific enzymes (60).

There are only a few experimental studies of hormonal effects which are generally relevant to the anxiety and avoidance learning problem as we have posed it. Mirsky *et al.* (60), using monkeys and rats, have made observations in three different learning situations, including avoidance procedures, and have obtained results consistent with the interpretation that adrenocorticotrophic hormone (ACTH) *decreases* anxiety or its drive properties. Using sheep, Liddell *et al.* (45) found that administration of adrenal extract resulted in a rapid *reduction* in tension and a disappearance of the rigidities and tic-like behavior which characterized "neurotic" sheep. These authors did not use an avoidance training procedure. However, in seeming contradiction to these results of Liddell and of Mirsky, Applezweig (7) has found that hypophysectomy in rats interferes with avoidance learning even though the capacity for escape learning continues. When these rats were treated with ACTH, they showed partial restoration of avoidance learning.

About all that can be said about experimental work in this area in its present preliminary stage is that the pituitary-adrenocortical system seems to have some sort of effect upon avoidance learning and presumably upon the secondary drive state mediating between CS and instrumental response; but the nature of such effects is obscure. It may be that the effects occur only under extreme experimental or pathological conditions, such as is the case with hypophysectomy.[3] Mirsky's investigations, inducing less extreme secondary physiological changes, deserve careful scrutiny because: (*a*) his results suggest that classically conditioned endocrine reactions may affect the course of avoidance learning by mechanisms which would be difficult to describe using current S-R or S-S formulations, and (*b*) because his results suggest that pituitary-adrenocortical reactions may counteract other kinds of classically conditioned reactions (negative, inhibitory, or inverse feedback?).

If endocrine reactions do, in fact, take part in classically conditioned anxiety, then the relatively long latencies of these reactions suggest that this portion of the total response may be strongly or inevitably subject to the principle of anxiety conservation by virtue of a prompt instrumental response. However, the tendency for such reactions to continue for long periods after the traumatic stimulus has been removed from the situation probably introduces a complication in the

[3] This may be unlikely. Recent studies of Gellhorn and his co-workers on insulin raise the possibility that hormonal mediating effects are powerful in maintaining conditioned responses. Gellhorn's most startling finding is that injections of insulin can result in the relatively permanent restoration of previously extinguished conditioned reactions (see Gellhorn, E. Is restoration of inhibited conditioned reactions by insulin coma specific for Pavlovian inhibitions? Contribution to the theory of shock treatment. *Arch. Neurol. Psychiat.*, 1946, **56**, 216–221).

applicability of this principle to endocrine reactions to a greater extent than it does to ANS reactions. Nevertheless, short exposure to the CS could be expected to minimize the magnitude of the endocrine reaction and, in that sense, to conserve it.

The structural changes in many parts of the body apparently brought by intense discharge of the adrenal cortex in the "adaptation syndrome" (73) of Selye may, particularly in the "exhaustion phase," be closely related to the concept of partial irreversibility as we have formulated it here. However, even if such reactions *can* be classically conditioned and are partially irreversible, their pertinence to avoidance learning theory is *not* great *unless* it can be demonstrated that they have behavioral effects.

Furthermore, even if such behavioral effects can be shown under certain special conditions, it will still remain to delineate the range of learning conditions in which such considerations are relevant. A preliminary piece of evidence in this unexplored area is the finding of Mirsky *et al.* (60) that ACTH has no effect upon avoidance responses which have been thoroughly established. In contrast, the same avoidance response at an earlier phase of learning readily undergoes extinction with the same amount of ACTH. This suggests that under these learning conditions, at least, the extent of the effect depends upon the *phase* of the learning process. Possibly other structures may take over later in the extinction phase.

Skeletal motor-proprioceptive system. In discussing the physiological aspects of avoidance learning from the viewpoint of the two-process theory, it is necessary to differentiate artificially and arbitrarily the effects of proprioceptive feedback in facilitating continuous postural adjustments from the functions which such feedback may serve as conditioned stimuli or as acquired drive stimuli. Schoenfeld (71), in a recent review, emphasizes the role of proprioceptive stimulation in serving as negative reinforcement (punishment) or as positive reinforcement (reward). He says: "The proprioceptive stimuli produced by the avoidance response may, because they are correlated with the termination of noxious stimuli, become secondary positive reinforcers and hence strengthen the tendency to make the response which generates them" (71, p. 88). Schoenfeld feels that the concept of proprioception is superior to that of emotion and fear in describing the data of avoidance learning.

Certainly the physiological and anatomical properties of proprioceptive feedback are more clear than for any other physiological system. However, we are ignorant of the relationships between particular kinds of proprioceptive events and behaviorial events in avoidance learning. There are always millions of proprioceptive impulses impinging on the CNS at any time. The measurement of those impulses which are of particular importance for the development and maintenance of learned avoidance responses is, therefore, an overwhelming problem. This is not so much the case for ANS and endocrine functions in which particular alterations may be experimentally created and their effects on avoidance learning studied. It is certainly premature to pass judgment on the relative utility of concepts like fear and anxiety and proprioception in describing the data of traumatic avoidance learning.

Light and Gantt (46) and Loucks (51) have shown that movement of a limb and the consequent proprioceptive stimulation are *not* necessary for the development of a skeletal CR. Lauer (42) has further shown that a conditioned neural discharge to the dog's limb muscles can be established under

total curarization. Thus, even when the dog was completely paralyzed, and distinctive proprioceptive correlates of muscular activity were therefore absent, avoidance responses were obtainable. These responses manifested themselves when the CS was presented after the effects of curarization had worn off. These observations do not rule out the possibility that proprioceptive feedback could facilitate, even though it is *not* essential to, the learning of avoidance.

It is well known that intense skeletal motor activity may lead to activation of the peripheral ANS and of many endocrine functions, with effects in the reverse direction also probable. Using dogs in which peripheral ANS function was eliminated before traumatic avoidance learning was begun, Wynne and Solomon (87) noted that these animals, compared to normal controls, showed much less motor "upset" and relatively little skeletal stereotyping of responses during extinction. This observation suggests that the ANS and its correlated feedback may have a more primary motivating role in such learning than does proprioceptive feedback. Perhaps the proprioceptive feedback may be inhibiting, contributing a work decrement factor.

Obviously, if the immediate motor reaction after presentation of a CS is an avoidance response, then other motor-proprioceptive processes which might be classified (more or less arbitrarily) as part of a total anxiety reaction, will not occur. In a sense, then, these other, later processes are conserved from extinction. This would apply to diffuse, classically conditioned skeletal responses and presumably not to other skeletal responses occurring at random. Schlosberg (70) has described the development of such skeletal responses.

Further analysis of some of these relationships may be possible using curare in dogs undergoing traumatic avoidance learning. An extremely important step in this direction has been taken in the ground-breaking experiment by Lauer (42), and we are currently following up this work by using curarized dogs in traumatic avoidance learning situations. The preliminary findings suggest that there is a rapid development of conditioned anxiety reactions in dogs under curare. However, the presence of such reactions does not necessarily result in rapid avoidance learning when the dog is later tested in normal condition. Rather, the skeletal effects of prior Pavlovian conditioning of anxiety under curare seem to be mostly composed of the diffuse postural adjustments and muscular tension emphasized by Schlosberg (70) in his account of classically conditioned skeletal reactions.

There seems to be no way of testing whether or not the motor-proprioceptive system may show partially irreversible changes in traumatic avoidance learning. If portions of this system in which partially irreversible structural changes might have developed are eliminated, the test response of instrumental avoidance is also eliminated. However, the rapidity of neural transmission plus the high degree of structural differentiation in this system make it seem a very unlikely locus of partial irreversibility.

Central nervous system. Within the CNS there are, clearly, numerous neural circuits which are aroused in massive pain-fear and anxiety reactions. Among these may be especially mentioned a well-defined circuit centering in the primitive forebrain which involves the hippocampus, fornix, mammillary body of the hypothalamus, anterior thalamic nuclei, limbic cortex, cingulum, and returns to the hippocampus (18). The primitive forebrain, loosely termed the "visceral brain" by MacLean (52), provides a system for integrating various intero- and exteroceptive impulses; complex autonomic functions are repre-

sented in a more organized manner here than in the circumscribed spot-to-spot representation of component autonomic functions in the premotor cortex (25, 53). Another important central network undoubtedly aroused in intense anxiety includes the connections between the reticular activating system of the lower brain stem, the thalamus, and the cerebral cortex (55). One or the other of these systems has been regarded by various authors (41, 47, 52, 67) as significant in the elaboration of "central emotion."

Direct physiological and anatomical evidence of connections with motor centers from primitive forebrain and reticular activating system is still incomplete. However, there do seem to be at least some association fibers from the primitive forebrain via the cingulate gyrus to cortical motor areas (85), and the reticular system apparently facilitates lower motor outflow (48). Major efferent discharge passes from the primitive forebrain via the hypothalamus to the peripheral ANS and to the pituitary-endocrine system (24, 53). Through such indirect mechanisms activity in these CNS association areas may affect skeletal motor responses in addition to the more direct CNS connections to motor areas.

Because of the difficulty of making direct measurements or observations of activity in such CNS circuits in response to a CS, these CNS reactions are awkward to describe as classically conditioned responses. Yet, if these CNS circuits can be regarded as the loci of "mediating" reactions which are activated by an external CS and which facilitate an instrumental CR, *then they are basically comparable to the peripheral acquired drives* described, for example, as occurring with ANS reactions.

Such an interpretation seems to be substantiated experimentally by the findings of Brady *et al.* (12). These authors have shown that cats receiving bilateral lesions of a portion of the primitive forebrain, namely, the amygdaloid complex and overlying cortex, acquired avoidance behavior of a given level with a significantly lower frequency than did operated or unoperated control animals. In this study the elimination of the ability to learn to avoid was incomplete and varied from animal to animal. This indicates that this part of the primitive forebrain, while not essential, probably is significant in the learning of avoidance. Whether a more complete elimination of the capacity to avoid would be obtained by ablation of other or greater portions of the primitive forebrain, or of other parts of the CNS, is still problematical.

Schreiner *et al.* (72), using the same cats as those in the avoidance learning study of Brady, observed marked changes in the sexual behavior of these animals which seem to be related to their endocrine functioning. These authors, therefore, suggest that the observed changes in behavior, precipitated by injury of the amygdaloid complex, may be due to a postsurgical state of altered endocrine activity. This possibility points up a difficulty in the interpretation of all ablation experiments in the study of avoidance learning, namely, that observed behavioral changes may be related to indirect and secondary changes produced by the ablation. However, the fact of the change suggests some sort of involvement of a specific brain area in the total sequence of physiological processes which underly the behavior being studied. A vast amount of work will be necessary, under standardized learning conditions, with various combinations of procedures which eliminate or alter various physiological functions, singly and in various combinations, before it will be possible to make full interpretations of the

physiological bases of such learning problems.

Little pertinent data are available concerning typical reaction latencies within these CNS circuits which have been thought to mediate "central anxiety." Part of the total latency, in the case of the reticular activating system, has been measured as 0.4 second, for the blocking of the alpha rhythm of the EEG after the onset of a sensory stimulus (47). Thus, these CNS latencies are probably discrete, though most likely a good deal shorter than for some of the peripheral circuits. Theoretically, such relatively rapid latencies in the CNS might reduce the applicability of the anxiety conservation principle. However, a prompt avoidance response would still reduce the duration of exposure to the CS and so would reduce the proportion of anxiety which is elicited and subject to ordinary extinction.

In our formulation of the principle of partial irreversibility of intense classically conditioned reactions, we have spoken of the phenomenon in terms of decreased threshold or sensitization. We have hypothesized that such changes would be especially likely to occur in those structures which are subjected to intense, convergent stimulation but which have a relatively primitive structure for differentiating such stimulation. Various authors have speculated concerning the possible nature of such changes in the CNS. MacLean believes that "it is possible that if a certain electrical pattern of information were to reverberate for a prolonged period or at repeated intervals in a neuronal circuit, the nerve cells (perhaps, say, as the result of enzymatic catalysis at specific axon-dendritic junctions) would be permanently 'sensitized' to respond to this particular pattern at some future time" (52, p. 349). Hebb has discussed the possibility that structural changes in the form of synaptic knobs may develop with intense or prolonged neuronal excitation (32).

Although the reticular activating system is probably subjected to a great variety of intense afferent stimulation under traumatic learning conditions, the structure of this system is apparently highly complex and differentiated. Therefore, we would not expect that the reticular system would be a major locus of partially irreversible changes. Because ablation of this system results in loss of consciousness of the organism, the possibility does not seem amenable to test using this experimental procedure. More promising is the primitive forebrain. This area, as we have already mentioned, brings into association a great variety of sensations, especially visceral feedback, which can be expected to be prominent in intense anxiety reactions. At the same time, the poorly differentiated cortical cytoarchitecture of the primitive forebrain (49, 50) suggests that it probably has little efficiency as an analyzer or discriminator of these convergent visceral and emotional functions.

These speculations are of especial interest because they seem to be in accord with actual experimental data of a preliminary sort. Brady *et al.* (12) found that cats which were given bilateral amygdala lesions *after* acquisition of avoidance showed a slight postoperative decrement in avoidance behavior. However, for the three animals in this experimental group, the decrement was not statistically significant. In the same study, Brady found that three other cats with lesions in the orbitalfrontal area, which is closely related to the primitive forebrain, showed a complete loss of previously well-established avoidance responses. After two days of postoperative retraining these animals

again reached the previous avoidance response level.

We have felt that partially irreversible changes in traumatic avoidance learning are related to the magnitude or intensity of convergent stimulation. Elimination of an appreciable portion of afferent feedback during acquisition of avoidance should interfere with the *development* of CNS partial irreversibility, but not affect partially irreversible changes *previously* established. The results of Wynne and Solomon (87) using procedures which eliminated peripheral ANS functioning are in accord with these inferences. They found that when these procedures were used *before* training was begun, avoidance behavior was much more easily extinguished than in normal control animals. However, if these procedures of ANS deprivation were carried out *after* avoidance had been learned with an intact ANS, then *no* effect upon the avoidance behavior was discernible. To be sure, these experiments did not determine what structures or processes had taken over in the latter instance; certainly CNS loci are most likely. Combinations of ablation experiments could test this possibility.[4] Thus, the physiological correlates of such hypotheses concerning avoidance behavior come to have empirically testable consequences which are intimately related to a potentially coherent theory of avoidance learning.

[4] The reader will probably note the absence of a discussion of the effects of frontal lobe ablations on anxiety reactions and avoidance responses. While such a discussion might be appropriate here, we felt that this would get us into special neurological problems which would take us too far afield. For example, it still is not clear to what extent the effects of frontal lobe ablations are an index of frontal lobe tissue damage, damage to connections between the frontal lobe, thalamus, and limbic system, or retrograde damage in lower brain centers themselves.

PROBLEMS RELATED TO LIFE HISTORY STAGES AND PSYCHOSOMATIC MEDICINE

The phenomenon of partial irreversibility is assumed to be *characteristic of classical conditioning only;* it is thought to result from massive feedback as a consequence of reactions to intense trauma. In thinking about the implications of such a principle, we felt that there might be some characteristics of infantile conditioning from which one could logically deduce some of the phenomena of early learning which various psychoanalysts, and in a different framework, Hebb, have discussed.

Freud (23) has maintained that early trauma produces long-lasting influences on behavior. In his conception of "primary anxiety," early traumatic experiences, mainly birth and infantile deprivation, are the basis from which later anxiety develops. Drawing upon a variety of clinical and experimental evidence, Greenacre (29) and Stern (82) have modified and elaborated earlier psychoanalytic formulations of this problem. Two of the properties of primary anxiety are a wide range of reactivating situations and a pervading enduringness of the tendency to reactivation. In contrast, postinfantile, situational anxiety has as two of its properties relative restriction to specific situations, and relative transientness. Hebb, approaching the same problem from the point of view of perceptual learning, has pointed out the enduring character of early perceptual organization: "Phase sequences" may be immutable through time when strongly established (32). We believe that a two-process learning theory, combined with the principle of partial irreversibility of classical conditioning, will deduce these phenomena directly.

In the first place, the infant is instrumentally helpless when contrasted

to the adult. Thus, there is a greater susceptibility to *intense* stimulation and a probability that specific CS-US sequences will be repeated many times. The infant may not have instrumental acts in its repertory with which it can terminate unconditioned stimuli quickly, or escape from conditioned stimuli. Thus, classical conditioning should be greatly facilitated in infantile stages. Pain-fear and anxiety reactions should be common, then, accompanied by diffuse skeletal reactions which are not particularly successful. This is not to state that the infant cannot control its environment to some extent. Rather, we think that the infant cannot do it as well as, say, the adolescent; and this leaves him, to a *relatively* great extent, a pawn of certain environmental stimulus sequences. In addition, early conditioning would be characterized by broader stimulus generalization, due mainly to the lack of conditioned discriminatory reactions and precise verbal symbols. Then, too, it is possible that a "new," possibly more "plastic" nervous system might be much more likely to show large-scale reorganization in the face of traumatic conditions. All of these considerations lead us to believe that early traumata are more likely to produce partial irreversibility of classical conditioning than would later traumata.

Later traumata will have somewhat different consequences. Here we can expect quick instrumental action, so that the repetition of CS-US pairings would be greatly cut down. We would expect a lot of conditioned avoidance reactions of a fairly discrete nature to develop, but no extremely strong classically conditioned anxiety reactions *in most cases*. With overwhelming trauma, however, we might find partial irreversibility with a single CS-US pairing, accompanied by avoidance responses of a specific sort. Clinical evidence suggests that this may happen with catastrophic war experiences, for example.

From one point of view, what we have suggested is that early, severe traumata are likely to produce classically conditioned emotional responses of a lasting sort. Later severe traumata are more likely to produce a variety of instrumental acts of a high strength. Visceral and hormonal conditioning should, therefore, be more characteristic of early traumata and be elicited by broadly generalized situational stimuli. Neurotic symptoms, such as phobic acts, compulsive behavior, etc., should be characteristic of later traumata, and they should have high specific symbolic content. If we accept such speculation, we then are led to conclude that many psychosomatic disturbances would be derivatives of early traumata, whereas the predominance of neurotic symptoms (avoidances, phobias, compulsions) should characterize later traumata.

One additional characteristic of later trauma should be considered. Verbal or symbolic capacity would greatly complicate stimulus generalization which takes place with reference to conditioned stimuli. Depending upon the attitudes of the individual, the prior categorizations of events, and degree of differentiations available at the verbal level, the meaning of a CS-US sequence could be quite complex. Reality testing would also be complex, therefore, and might be carried on symbolically to a great extent. This complex differentiation would be less the case in instances of infantile traumata. Stimulus generalization would follow inherent gradients, would be relatively free of specific symbolic interpretations. Reality testing would be more closely bound to extinction principles, and would consist mainly of dissociating conditioned stimuli from unconditioned stimuli. Thus, we would expect reality

testing in later traumata to be highly verbal; reality testing for early traumata would be highly inarticulate.

RELATED PSYCHOTHERAPEUTIC PROBLEMS

It is interesting to note some of the implications of the principles of anxiety conservation and partial irreversibility for the appropriateness of psychotherapeutic procedures for various kinds of disorders. Psychotherapy relying largely upon the environmental manipulation and alteration of obsessive, compulsive, or phobic behavior would, using our formulation, seem to depend upon the substitution of one instrumental response for another. Such therapeutic procedures might bring about more socially desirable behavior, but could not be expected to remove completely the classically conditioned anxiety reactions which elicit one form or other of the neurotic behavior.

Another therapeutic approach might involve the introduction of competing motivation or the introduction of stimulus situations which elicit strong responses that are incompatible with neurotic avoidance responses. But the individual undergoing such therapeutic procedures would still "feel uncomfortable or anxious" in certain situations. This would be true even if the whole cognitive organization of the individual had been drastically modified. In a sense, all the various cognitive and manipulative procedures can be regarded as conserving anxiety from extinction processes. For example, compulsive hand washing is typically carried out before anxiety over real or fantasied dirt is fully experienced. When such patients are prevented from washing their hands under such stimulus conditions, marked anxiety or panic may ensue. From this standpoint, the skilled psychotherapist is able to help the patient to let his anxiety be re-experienced in a supportive setting (so that the therapist is not thereafter avoided in common with other panic-provoking stimuli!).

Such a therapeutic procedure is in accord with the principle that classically conditioned reactions must actually occur if they are to undergo extinction. A considerable decrease in the intensity of symptoms can be expected if the therapist establishes a relationship with the patient which duplicates parts of the conditioned stimulus pattern associated with original traumatic experiences. Optimally, an intense transference relationship can make possible nonverbal reality testing and the *partial* alleviation of anxiety having preverbal origins.

However, according to the principle of the partial irreversibility of traumatic anxiety reactions, there will be certain definite limitations on the "curing" of behavior arising from early, "primitive" traumatic experiences. This will also hold true for psychosomatic symptoms which may be a more direct manifestation of early conditioning. Complete freedom from a *tendency* to manifest such symptoms could not be expected, even with the most advantageous course of therapy.

(The authors have been informed by several individuals who have personally undergone prolonged and successful psychoanalysis that conscious efforts to relive in fantasy their most anxiety-provoking life experiences still reactivate a discernible residual of this anxiety which was once much more considerable in degree. However, these individuals feel they have sufficiently learned to discriminate these anxiety-provoking situations of early life from those situations ordinarily occurring in adulthood. Such distinctions, of course, involve gradients, not sharp demarcations.)

SUMMARY

We have presented a highly speculative analysis of anxiety and avoidance learning. The central assumptions of the analysis were: (*a*) there are two basic acquisition processes, the classical conditioning of emotional reactions and the instrumental learning of skeletal responses; (*b*) the laws of emotional conditioning are those of Pavlovian conditioning, where the relationship between a CS and a US is of prime importance; (*c*) the laws of skeletal response modification are those of reward and punishment learning, a Thorndikian or Hullian conception; (*d*) in the case of avoidance conditioning, the analysis of extinction phenomena can be more easily made if we introduce the principle of "anxiety conservation," a theoretical label for complex interrelations between the strength of anxiety reactions and instrumental avoidance responses; the central idea in the principle of anxiety conservation is that short-latency avoidance responses will prevent the CS from arousing anxiety reactions, thereby conserving conditioned anxiety reactions from extinction; (*e*) in the case of intense trauma as the US, the classical conditioning of emotional reactions is partially irreversible, and extinction of such reactions can only be partial at best; the central idea in the "principle of partial irreversibility" of emotional conditioning is that feedback from peripheral ANS reactions results in an overloading of rather primitively differentiated areas of the "emotional brain."

The consequences of such an analysis have been discussed with an eye to both behavioral and physiological research on anxiety and avoidance learning. We have included highly selected applications of our analysis to research on avoidance learning, psychotherapy, and psychosomatic medicine. We have specifically worked out implications of the analysis for: (*a*) extinction of anxiety reactions, (*b*) extinction of instrumental avoidance responses, (*c*) the effectiveness of reality testing therapeutic procedures, (*d*) the use of reward and punishment in facilitating the extinction of instrumental avoidance responses, (*e*) the effects of trauma at various stages in life history, and (*f*) the limitations and appropriateness of certain psychotherapeutic methods in alleviating anxiety and overt neurotic behavior sequences. We have not hesitated to introduce very speculative physiological considerations wherever we felt that it might stimulate new research or suggest new approaches. The role of the peripheral ANS, the endocrine system, the proprioceptive system, and certain areas of the CNS were discussed relative to anxiety reactions and avoidance learning.

In conclusion, the authors feel that they have not enunciated a "theory." Rather, they have presented some ideas and speculations which have been personally useful in organizing some facts of psychology. Some of these ideas have empirical consequences and so the wildness of speculation may perhaps be excused.

REFERENCES

1. AIDAR, O., GEOHEGAN, W. A., & UNGEWITTER, L. H. Splanchnic afferent pathways in the central nervous system. *J. Neurophysiol.*, 1952, **15**, 131–138.
2. ALLEN, W. F. Studies on irradiated cerebral differentiated excitation and inhibition as indicated and measured by respiration. *Amer. J. Physiol.*, 1942, **136**, 783–795.
3. ALLPORT, G. W. *Personality: a psychological interpretation.* New York: Holt, 1937.
4. AMASSIAN, V. E. Cortical representation of visceral afferents. *J. Neurophysiol.*, 1951, **14**, 433–444.
5. AMASSIAN, V. E. Fiber groups and spiral pathways of cortically represented visceral afferents. *J. Neurophysiol.*, 1951, **14**, 445–460.

6. AMASSIAN, V. E. Interaction in the somatovisceral projection. *Res. Publ. Ass. nerv. ment. Dis.*, 1952, **30**, 371–402.
7. APPLEZWEIG, M. H., & MOELLER, G. Hormonal influences in learning: the pituitary-adrenal system, anxiety, and avoidance learning. ONR project reports NR 154–103 and NR 154–137, 1953.
8. AULD, F., JR. The effects of TEA on a habit motivated by fear. *J. comp. physiol. Psychol.*, 1951, **44**, 565–574.
9. BAILEY, P., & BREMER, F. A sensory cortical representation of the vagus nerve. *J. Neurophysiol.*, 1938, **1**, 405–412.
10. BITTERMAN, M. E., REED, P., & KRAUSKOPF, J. The effect of the duration of the unconditioned stimulus upon conditioning and extinction. *Amer. J. Psychol.*, 1952, **45**, 256–262.
11. BRADY, J. V. Does tetraethylammonium reduce fear? *J. comp. physiol. Psychol.*, 1953, **46**, 307–310.
12. BRADY, J. V., SCHREINER, L., GELLER, I., & KLING, A. Subcortical mechanisms in emotional behavior: the effect of rhinencephalic injury upon the acquisition and retention of a conditioned avoidance response in cats. *J. comp. physiol. Psychol.*, 1954, **47**, 179–186.
13. BROGDEN, W. J. Acquisition and extinction of a conditioned avoidance response in dogs. *J. comp. physiol. Psychol.*, 1949, **42**, 296–302.
14. BROWN, J. S., & JACOBS, A. The role of fear in the motivation and acquisition of responses. *J. exp. Psychol.*, 1949, **39**, 747–759.
15. BRUSH, F. R., BRUSH, ELINOR S., & SOLOMON, R. L. Traumatic avoidance learning: the effect of the CS-US interval on acquisition and extinction. Paper read at East. Psychol. Ass., Boston, Mass., April, 1953.
16. CANNON, W. B. *The wisdom of the body.* New York: Norton, 1932.
17. CLARK, L. D., QUARTON, G. C., COBB, S., & BAUER, W. Further observations on mental disturbances associated with cortisone and ACTH therapy. *New Eng. J. Med.*, 1953, **249**, 178–183.
18. CLARK, W. E. LE GROS, & MEYER, M. Relationships between the cerebral cortex and the hypothalamus. *Brit. Med. Bull.*, 1950, **6**, 341–344.
19. CLEGHORN, R. A. Endocrine influence on personality and behavior. In *The biology of mental health and disease.* New York: Hoeber, 1952. Pp. 265–276.
20. COBB, S. *Emotions and clinical medicine.* New York: Norton, 1950.
21. DYKMAN, R. A., & GANTT, W. H. A comparative study of cardiac and motor conditional responses. *Amer. J. Physiol.*, 1951, **167**, 780.
22. FRENCH, J. D., VON AMERONGEN, F. K., & MAGOUN, H. W. An activating system in brain stem of monkey. *Arch. Neurol. Psychiat.*, 1952, **68**, 577–590.
23. FREUD, S. *The problem of anxiety.* New York: Norton, 1936.
24. FULTON, J. F. *Physiology of the nervous system.* (3rd Ed.) New York: Oxford Univer. Press, 1949.
25. FULTON, J., PRIBRAM, K. H., STEVENSON, J. A. F., & WALL, P. Interrelations between orbital gyrus, insula, temporal tip, and anterior cingulate. *Trans. Amer. Neurol. Ass.*, 1949, **74**, 175–179.
26. GANTT, W. H. *Experimental basis for neurotic behavior.* New York: Hoeber, 1944.
27. GANTT, W. H. Principles of nervous breakdown—schizokinesis and autokinesis. *Ann. N. Y. Acad. Sci.*, 1953, **56**, 143–163.
28. GRAHAM, B. F. Neuroendocrine components in the physiological response to stress. *Ann. N. Y. Acad. Sci.*, 1953, **56**, 184–199.
29. GREENACRE, P. *Trauma, growth, and personality.* New York: Norton, 1952.
30. GWINN, G. T. The effects of punishment on acts motivated by fear. *J. exp. Psychol.*, 1949, **39**, 260–269.
31. HARRIS, G. W. The hypothalamus and endocrine glands. *Brit. Med. Bull.*, 1950, **6**, 345–350.
32. HEBB, D. O. *Organization of behavior.* New York: Wiley, 1949.
33. HILGARD, E. R., & MARQUIS, D. G. *Conditioning and learning.* New York: D. Appleton-Century, 1940.
34. HULL, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
35. HUME, D. M., & WITTENSTEIN, G. J. The relationship of the hypothalamus to pituitary-adrenocortical function. In J. R. Mote (Ed.), *Proceedings of 1st Clinical ACTH Conference.* Philadelphia: Blakiston, 1950.
36. JENKINS, W. O., & STANLEY, J. C. Partial reinforcement: a review and critique. *Psychol. Bull.*, 1950, **47**, 193–234.
37. JERSILD, A. T., & HOLMES, F. B. Methods of overcoming children's fears. *J. Psychol.*, 1935, **1**, 75–104.

38. JONES, H. E. The retention of conditioned emotional responses in infancy. *J. genet. Psychol.*, 1930, **37**, 485–498.

39. KAMIN, L. J. Traumatic avoidance learning: the effects of CS-US interval with a trace conditioning procedure. *J. comp. physiol. Psychol.*, 1954, **47**, 65–72.

40. KIMBLE, G. A. A description of avoidance learning. *Amer. Psychologist*, 1952, **7**, 271. (Abstract)

41. KUBIE, L. S. Some implications for psychoanalysis of modern concepts of the organization of the brain. *Psychoanal. Quart.*, 1953, **22**, 21–68.

42. LAUER, D. W. Conditioning under complete curarization. *Amer. Psychologist*, 1951, **6**, 280. (Abstract)

43. LICHTENSTEIN, P. E. Studies of anxiety: I. The production of a feeding inhibition in dogs. *J. comp. physiol. Psychol.*, 1950, **43**, 16–29.

44. LIDDELL, H. S. Conditioned reflex method and experimental neurosis. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald, 1944. Pp. 389–412.

45. LIDDELL, H. S., ANDERSON, O. D., KOTYUKA, E., & HARTMEN, F. A. Effect of extract of adrenal cortex on experimental neurosis in sheep. *Arch. Neurol. Psychiat.*, 1935, **34**, 973–993.

46. LIGHT, J. S., & GANTT, W. H. Essential part of reflex arc for establishment of conditioned reflex; formation of conditioned reflex after exclusion of motor peripheral end. *J. comp. Psychol.*, 1936, **21**, 19–36.

47. LINDSLEY, D. B. Emotion. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 473–516.

48. LINDSLEY, D. B. Brain stem influences on spinal motor activity. *Res. Publ. Ass. nerv. Ment. Dis.*, 1952, **30**, 174–195.

49. LORENTÉ DE NO, R. Studies on the structure of the cerebral cortex: I. The area entorhinalis. *J. Psychol. Neurol., Lpz.*, 1933, **45**, 381–437.

50. LORENTÉ DE NO, R. Studies on the structure of the cerebral cortex: II. Continuation of the study of the ammonic system. *J. Psychol. Neurol., Lpz.*, 1934, **46**, 113–177.

51. LOUCKS, R. B. The experimental delimitation of neural structures essential for learning: the attempt to condition striped muscle responses with faradization of the sigmoid gyri. *J. Psychol.*, 1935–36, **1**, 5–44.

52. MACLEAN, P. D. Psychosomatic disease and the "visceral brain": recent developments bearing on the Papez theory of emotions. *Psychosom. Med.*, 1949, **11**, 338–351.

53. MACLEAN, P. D. Some psychiatric implications of physiological studies on frontotemporal portion of limbic system (visceral brain). *EEG clin. Neurophysiol.*, 1952, **4**, 407–418.

54. MAGOUN, H. W. An ascending reticular activating system in the brain stem. *Arch. Neurol. Psychiat.*, 1952, **67**, 145–154.

55. MAIER, N. R. F. *Frustration, the study of behavior without a goal*. New York: McGraw-Hill, 1949.

56. MASSERMAN, J. H. *Behavior and neurosis*. Chicago: Univer. of Chicago Press, 1943.

57. MASSERMAN, J. H., & PECHTEL, C. Neuroses in monkeys: a preliminary report of experimental observations. *Ann. N. Y. Acad. Sci.*, 1953, **56**, 253–265.

58. MILLER, N. E. Studies of fear as an acquirable drive: I. Fear as motivation and fear-reduction as reinforcement in the learning of new responses. *J. exp. Psychol.*, 1948, **38**, 89–101.

59. MILLER, N. E. Learnable drives and rewards. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 435–472.

60. MIRSKY, I. A., MILLER, R., & STEIN, M. Relation of adrenal cortical activity and adaptive behavior. *Psychosom. Med.*, in press.

61. MOWRER, O. H. Anxiety-reduction and learning. *J. exp. Psychol.*, 1940, **27**, 407–516.

62. MOWRER, O. H. On the dual nature of learning—a reinterpretation of "conditioning" and "problem-solving." *Harv. educ. Rev.*, Spring, 1947, 102–148.

63. MOWRER, O. H. *Learning theory and personality dynamics*. New York: Ronald, 1950.

64. MOWRER, O. H., & JONES, H. M. Habit strength as a function of the pattern of reinforcement. *J. exp. Psychol.*, 1945, **35**, 293–311.

65. MOWRER, O. H., & LAMOREAUX, R. R. Avoidance conditioning and signal duration—a study of secondary motivation and reward. *Psychol. Monogr.*, 1942, **54**, No. 5 (Whole No. 247).

66. MOWRER, O. H., & LAMOREAUX, R. R. Fear as an intervening variable in avoidance conditioning. *J. comp. Psychol.*, 1946, **39**, 29–50.

67. PAPEZ, J. W. A proposed mechanism of emotion. *Arch. Neurol. Psychiat.*, 1937, **38**, 725.
68. PAVLOV, I. P. *Conditioned reflexes.* London: Oxford Univer. Press, 1927.
69. ROME, H. P., & BRACELAND, F. J. Use of cortisone and ACTH in certain diseases: psychiatric aspects. *Proc. Staff Meet., Mayo Clinic*, 1950, **25**, 495–497.
70. SCHLOSBERG, H. The relationship between success and the laws of conditioning. *Psychol. Rev.*, 1937, **44**, 379–394.
71. SCHOENFELD, W. N. An experimental approach to anxiety, escape and avoidance behavior. In P. H. Hoch & J. Zubin (Eds.), *Anxiety*. New York: Grune and Stratton, 1950. Pp. 70–99.
72. SCHREINER, L., & KLING, A. Behavioral changes following rhinencephalic injury in the cat. *J. Neurophysiol.*, in press.
73. SELYE, HANS. *The physiology and pathology of exposure to stress.* Montreal: Acta, Inc., 1950.
74. SHEFFIELD, F. D. Avoidance training and the contiguity principle. *J. comp. physiol. Psychol.*, 1948, **41**, 165–177.
75. SHEFFIELD, F. D., & TEMMER, H. W. Relative resistance to extinction of escape training and avoidance training. *J. exp. Psychol.*, 1950, **40**, 287–297.
76. SHEFFIELD, VIRGINIA F. Extinction as a function of partial reinforcement and distribution of practice. *J. exp. Psychol.*, 1949, **39**, 511–525.
77. SKINNER, B. F. Two types of conditioned reflex and a pseudo type. *J. gen. Psychol.*, 1935, **12**, 66–77.
78. SOLOMON, R. L., KAMIN, L. J., & WYNNE, L. C. Traumatic avoidance learning: the outcomes of several extinction procedures with dogs. *J. abnorm. soc. Psychol.*, 1953, **48**, 291–302.
79. SOLOMON, R. L., & WYNNE, L. C. Avoidance conditioning in normal dogs and in dogs deprived of normal autonomic functioning. *Amer. Psychologist*, 1950, **5**, 264. (Abstract)
80. SOLOMON, R. L., & WYNNE, L. C. Traumatic avoidance learning: acquisition in normal dogs. *Psychol. Monogr.*, 1953, **67**, No. 4 (Whole No. 354).
81. STARZL, T. E., & WHITLOCK, D. G. Diffuse thalamic projection system in monkey. *J. Neurophysiol.*, 1952, **15**, 449–468.
82. STERN, M. M. Anxiety, trauma, and shock. *Psychoanal. Quart.*, 1951, **20**, 179–203.
83. TRETHOWAN, W. H., & COBB, S. Neuropsychiatric aspects of Cushing's syndrome. *Arch. Neurol. Psychiat.*, 1952, **67**, 283–309.
84. WALL, P. D., & DAVIS, G. D. Three cerebral cortical systems affecting autonomic function. *J. Neurophysiol.*, 1951, **14**, 507–517.
85. WARD, A. A., JR. The anterior cingulate gyrus and personality. *Res. Publ. Ass. nerv. ment. Dis.*, 1948, **27**, 438–445.
86. WHITE, R. W. *The abnormal personality.* New York: Ronald, 1948.
87. WYNNE, L. C., & SOLOMON, R. L. Traumatic avoidance learning: acquisition and extinction in dogs deprived of normal peripheral autonomic function. *Genet. Psychol. Monogr.*, in press.
88. ZANCHETTI, A., WANG, S. C., & MORUZZI, G. The effect of vagal afferent stimulation on the EEG pattern of the cat. *EEG clin. Neurophysiol.*, 1952, **4**, 357–361.
89. ZENER, K., & MCCURDY, H. G. Analysis of motivational factors in conditioned behavior: I. The differential effect of changes in hunger upon conditioned, unconditioned, and spontaneous salivary secretion. *J. Psychol.*, 1939, **8**, 321–350.

(Received February 4, 1954)

Psychological Review
Vol. 61, No. 6, 1954

THE PSYCHOANALYTIC THEORY OF CONFLICT: STRUCTURE AND METHODOLOGY [1]

ALLEN T. DITTMANN [2] AND HAROLD L. RAUSH
University of Michigan

Although some criticisms of psychoanalytic theory can be laid at the door of the personal motivations and unconscious resistances of the critics, there is something to the view that workers in the field do not have available to them any concise and complete statement of the structure of psychoanalytic theory. Accordingly, there is a large group who mistake the *content* of psychoanalytic findings for the *structure* of psychoanalytic theory, and believe, for example, that the stages of psychosexual development and the standard list of defense mechanisms are coterminous with the theory itself. Caught in this confusion, many investigators have contented themselves with attempting to "prove" or "disprove" "psychoanalysis" on the basis of studies of certain content elements of the theory. There can be no doubt that the investigation of isolated aspects of psychoanalytic theory may furnish us with factual information and with broad differentiations. We believe, however, that only through understanding and utilizing the structural interrelations among the concepts of the theory can we make use of the real power and research potential which the theory may offer, or enter into legitimate conceptual refinements and revisions.

The purpose of this paper is, therefore, to outline these structural interrelations. We shall concern ourselves with the psychological *processes* that psychoanalysis posits as necessary for the understanding of certain behaviors. In view of our limited purpose, it will be necessary to omit or reduce discussion of many relevant aspects and complexities of psychoanalytic theory. Thus we shall not deal, for example, with the theories of psychosexual stages, system structure (id, ego, superego), intrasystem problems, differentiation of impulses (sex and aggression), primary and secondary processes, or with many clinical and technical problems. Our presentation is drawn from Freudian psychoanalytic theory; after presenting the theoretical outline, we shall examine the research methodology from which this viewpoint developed, and discuss some research directions toward which psychoanalysis points.

Sources of Data for the Theory

In order to understand the structure of the theory and its sometimes confusing syntax, it is first necessary to look at the sources of data used in its development. Many sources have been drawn upon, from ethnological field reports to children's drawings, but all of them have been used chiefly as collaborative evidence for an already growing theory: the primary source of data has always been the productions of psychoneurotic patients undergoing therapeutic analyses. The fact that the data supplied by patients did not fit contemporary notions led Freud to break away from his late nineteenth century coworkers and to start the development of psychoanalytic therapy. Similarly, the fact that the data of the treatment

[1] The writers wish to express appreciation to Edward S. Bordin, Donald G. Marquis, and Frederick Wyatt for their critical reading of the manuscript. The responsibility for the views expressed is, however, that of the two authors, who shared equally in the development and writing of the paper.

[2] Now at National Institute of Mental Health, Bethesda, Maryland.

situation did not fit his own early formulations led Freud and his new coworkers to make every major change, expansion, and development that the theory has undergone up to the present time.

Psychoanalysis developed in a clinical setting as a therapeutic measure, at first particularly for hysterical patients; it was primarily a psychology of the unconscious (13, p. 128). One of the first clinical facts that Freud recognized was that a psychology not based on unconscious sources of motivation was inadequate either for understanding or for treating neurotic patients. The method of free association was gradually evolved as the most economical method for clarifying unconscious sources of behavior and their effects. The basic assumption of this method is that free associations are indeed not "free," but are rather highly determined by unconscious motivations. The fundamental rule "to say everything that comes into your mind, whether it seems important or unimportant, relevant or irrelevant," is designed to minimize the influence of consciously directed thought processes, and to maximize the influence of unconscious factors. Other aspects of the analytic situation as it is usually constituted—the reclining position of the patient, the subdued atmosphere of the consulting room, the position of the analyst out of direct sight of the patient—are similarly designed to maximize the influence of unconscious factors in the patient's "free" associations.

Derivatives as behavioral resultants of unconscious conflict. Under these traditionally analytic conditions, many, though not necessarily all, of the patient's productions in the analytic hour are derivatives of his unconscious conflicts. That is, these productions are representatives of the same conflicts that have led him into the difficulties for which he has sought treatment. Precisely what conflict the patient's productions represent is at first unknown to both patient and analyst. Even the fact that his productions do represent conflict may be unknown to or of no concern to the patient, but this fact is the major assumption under which the analyst operates. Were the associations (or sometimes the lack of them) not representative of conflict, and were they not *disguised* representatives of *unconscious* conflict (that is, derivatives), the patient would have little need or motivation for treatment—he could deal with reality as adequately as the situation allowed, and with a minimal degree of discomfort.

Let us consider a clinical example of the production of derivatives, an example which we may use for further discussion of derivatives and of their importance to the structure of psychoanalytic theory. A student has difficulty in submitting papers on time. While he feels that he is capable of "A" work, he usually manages to get an A − or a B +. Problems such as this are so common in student populations that to speak of them as derivatives of unconscious conflicts is at this point foolhardy. But let us go a step further. When our student gets an A, we are surprised to learn that he feels vaguely dissatisfied: the instructor has misunderstood him, has praised the wrong things. What is more, our student now feels that the ideas are no longer his, that he has given in to external pressure and has let his ideas and thoughts be taken away from him. Yet at the same time he does want to express these ideas and wants them to be appreciated. His communications to the therapist follow the same line. As soon as he has communicated a thought, he feels dissatisfied and tries in various ways to nullify what he has said. He feels that the thoughts he has expressed are no longer his, but are rather the property

of the therapist. He blocks at a point where he is discussing his difficulties in communicating to the therapist, and the latter inquires as to the thought at the point of blocking. The patient embarrassedly confesses having thought of feces, but disclaims this as an irrelevant, nasty thought, and can go no further with the subject. We continue to discover similar patterns in his other interpersonal relationships: we find that he wants to give himself up completely in an all-encompassing intimacy with his friends, surrendering his ideas and his will, yet he finds himself in sudden bouts of temper and resentment, fighting in strange ways to maintain his integrity. He sees women clutching at him and trying in devious ways to trap him and get something out of him. Sexual expression and sexual curiosity he feels are signs of his vulnerability and must be denied. At times he fantasies being overwhelmed and raped by one more powerful than he.

THE THEORETICAL STRUCTURE OF PSYCHOANALYSIS

The derivative as compromise behavior. In clinical examples such as the one cited above, the therapeutic situation has allowed for intensive, long-time exploration of the meaning of the derivatives in terms of conflicts underlying them. We have the opportunity to learn how such seemingly superficial behaviors as not submitting term papers on time are ramifications of early childhood experiences. We may, however, observe the formation of derivatives in other situations, too, although the opportunities for exploring their sources are usually not so available to us. Dreams, slips of the tongue, symptoms are examples. All of these phenomena are compromises between opposing forces which make up unconscious conflicts within individuals. The behavior involved in the compromise always expresses in a disguised form some impulse which would be otherwise unacceptable. The disguise itself functions so as to allow expression of the impulse in such a way that the anxiety which would follow direct expression of the impulse is not so likely to be aroused. The processes by which the disguises are accomplished are known as the defensive functions of the ego. Thus, the very naming of a bit of behavior as a derivative implies a complex theoretical structure: an impulse striving for expression, the direct expression of which would involve anxiety, and defensive maneuvers to allow some expression with minimal anxiety.[3]

Impulse

Let us begin our analysis of this theoretical structure with a consideration of impulses. The exact nature of basic impulses is not known, and the issue of their explanation is a confused one. Freud, himself, seems always to have been dissatisfied with his formulations in this area, and was careful to point out the tentative state of his conceptualizations.[4]

[3] Psychoanalytic writers using the term "derivative," generally speak of "derivatives of unconscious impulses." These refer, as above, to distorted forms of impulse expression which occur when direct impulse expression is impossible because of anxiety. The context of the writings makes it clear that the distortion process is as important as the expression, and that derivatives may be considered in terms of their degree of distortion of the original impulse. Since the distortion is a function of defensive processes, the derivative always represents both impulse and defense (6, p. 57). We believe, then, that it is more consistent to speak of a derivative as the resultant of an impulse-anxiety-defense conflict.

[4] One such remark was made in the *New Introductory Lectures on Psycho-analysis:* "The theory of the instincts is, as it were, our mythology" (15, p. 131). The historical development and problems associated with this theory are discussed by Bibring (2).

We use the English word "impulse" here for

We shall tentatively define impulse as a physiologically based tension state that sets behavior in motion. Freud (11) describes impulses as having (*a*) impetus, which is the motor element of drive, the force it represents; (*b*) aim, which is, in general, discharge of the energy or satisfaction; (*c*) object, which is the instrument through which satisfaction may be attained; and (*d*) source, which is the physicochemical state of the organism.[5]

Impetus. We may consider impetus as representing the quantitative strength of an impulse, a strength which varies from time to time. It decreases when certain specific behavior patterns occur in conjunction with specific environmental opportunities; it increases either where the specific behavior pattern fails to occur or where the environment fails to present the appropriate opportunities. In words more generally used by psychoanalysts, the impetus of an impulse decreases with satisfaction and increases with inhibition or deprivation.

Aim. While it is true that the aim of impulses is always toward discharge, toward an eventual lowering of tension level, the process may be inhibited or deflected, and the modes of achieving this aim may be varied. At times Freud seems to use the word "aim" to include differing modes of impulse satisfaction. An example of this confusion may be found in his discussion of the "change of aim" of an impulse from active to passive, as scoptophilia to exhibitionism, and sadism to masochism (11, p. 72). Here the phrase "change of aim" should more properly read "change of mode for achieving aim," as Freud wrote elsewhere in the same paper (11, p. 69).

It should be noted that the immediate satisfaction of some impulses, such as respiration, is imperative to the life of the organism, whereas for other impulses, greater and perhaps indefinite delay can be tolerated. Similarly modes of satisfaction are more variable for some impulses than for others.

Object. Impulses vary in the range of objects that may serve to reduce their impetus. Respiratory needs, for example, can be satisfied only by breathing air. Sexual impulses, on the other hand, can be satisfied by a wide variety of objects and by various modes. The impulses most relevant to psychology are those most amenable to variation in mode of expression and in object. These impulses are important because they are the ones most subject to environmental influences and hence can become the foci of interpersonal conflict.

Source. Freud's belief in physicochemical states as the sources of all impulses forms the basis of his biological viewpoint. He even hoped for a state of knowledge which would allow classification of impulses on the basis of their sources (11). We are still unable to fulfill these hopes, although the time may yet come when we will be able to do so.

Psychological Representations of Impulse

In talking about impulses as physiological states, it is of the utmost importance to recognize that these states themselves are not psychological phenomena. The psychological phenomena are, rather, the tensions experienced by the individual, tensions which carry with them, to a greater or lesser degree, content of associations with other ex-

the German *Trieb*, which has usually been translated "instinct," and occasionally "drive" (7, p. 12).

[5] Freud appears to have dropped the term impetus in his later writings, though equivalent concepts appear in his discussion of strength of instincts (16); Fenichel (7) discusses aim, object, and source as characteristics of instincts, but Sterba (25) includes the concept of impetus in his discussion, though he calls it "drive."

periences. These tensions are not necessarily related in a one-to-one fashion with fluctuations in the physiological state of the organism. Thus, operationally it is necessary to distinguish between the physiological state that results from food deprivation, for example, and the experience of the resulting tensions. This psychological hunger varies not only as a function of the physiological state, but also as a function of such conditions as age and previous experience of the individual. These psychological concomitants of physiological states are what Freud called "impulse presentations" (12). In most clinical writing the concepts of impulse and impulse presentation have unfortunately been confused, so that the term impulse has come to be used as a standard expression for impulse presentation. Consequently, in psychoanalytic literature, impulse variously refers to: (*a*) an energic, "driving," physiological force; (*b*) the psychological representations of such forces; and (*c*) residues of experiences with such forces.

We may illustrate this confusion by considering again the case of the student cited above. For this man, current sexual tensions, when they arise, are accompanied by a multiplicity of contents. It may be possible to speak of each of these contents as independent impulses or as fusions of them, e.g., scoptophilic, passive-dependent, anal, homosexual, and other impulses. However, to the extent that we believe these factors to be modifiable, an inconsistency is involved in speaking of them as impulses in the sense of independent physiological tension states. We should rather speak of such tendencies as impulse presentations concomitant with a physiological imbalance, which in this case is sexual tension. It seems reasonable to suppose that our patient's twistings and turnings with respect to sexual impulses are not a function of struggle between impulses, but rather that they are a function of memory traces (cf. 9, ch. VII) resulting from his own unique experiences. These experiences have influenced the course of handling states of physiological imbalance.

In the case of another patient, psychological tensions were always so high that no one physiological state could be differentiated from any other. She stated, for example, that she never "felt hungry," but ate when her watch pointed to the traditional times for meals. In the course of treatment she came to know periods of relaxation during which specific tensions could be felt. Impulse and impulse presentation were remarkably separate from each other for this patient.

Anxiety

Primary anxiety or panic. For any individual, experiences may have been favorable or unfavorable to the expression of a given impulse, that is, to the correct interpretation of a physiological imbalance and to behavior that will lead to the reduction of this imbalance. Where experience is favorable, the individual gradually develops techniques for achieving impulse satisfaction, and evaluations of the appropriate times and places for impulse satisfaction. Such learnings may be made simple or complex, either by society or by parents as particular representatives of society. As long as society provides channels for discharge and adequate rewards for approved modes of discharge, learning is a continuously modifying process. That is, experience becomes checked and modified by new experience.

But in the case of our patient, we may note that he avoids those situations which are appropriate to impulse expression, and that he misinterprets

both the external world and what we infer to be his physiological needs. He is furthermore strikingly persistent in his misinterpretations, despite the seeming discomfort that these misinterpretations bring. The evidence he presents leads us to believe that experiences in his life had the peculiar effect of preventing him from profiting by new experiences. How does psychoanalytic theory account for such a state of affairs?

The expression of any impulse or of all impulses may be accompanied by intense threat in the early experience of an individual. This threat consists of the overwhelming of the psychic apparatus by tensions attendant upon physiological imbalances that the infant cannot satisfy by himself. These imbalances may be the result of the normal periodic rise of needs which parental figures have not yet satisfied at the moment, or they may result from external causes such as parental treatment, disease, and the like. Where parental treatment in relation to impulse expression leads to such overwhelming threat, the threat is likely to be a repeated experience, either because of institutionalized child-rearing practices or because of the individual parent's persistent needs and fears. The response of the infant to such trauma has been called primary anxiety (7, pp. 132 ff.).

In its theory of genetic development, psychoanalysis proposes certain physiological and psychological stages. For the purposes of the present paper it is not necessary to go into the details of these stages; they have been described adequately elsewhere (3, 5, 7). What is relevant here is that these stages are associated with certain impulses and threats to their expression, and consequently have potentiality for becoming focal points for primary anxiety.[6] The panic experience of primary anxiety is so painful that the psychological representation of the impulse comes to be regarded as dangerous. Similarly, affects and images which are concomitant with the experience of primary anxiety often come to be perceived as dangerous. According to the theory, such experiences and their psychological connections tend to undergo repression and are retained as unconscious memory traces.

Secondary anxiety and memory traces. Having experienced primary anxiety, the infant is sensitized to those events which might rearouse panic. These events may take two forms: (*a*) from the impulse side, any increase in the impetus of the specific impulse; (*b*) from the side of the external world, any situation which is provocative of impulse expression or which is similar to the events surrounding the earlier experience of panic. In either form, such events bring about expectations on the part of the individual that a panic of the same intensity as the original one will arise. The individual is consequently always on the alert for cues from within and from without that will signal an impending danger similar to that experienced in the past. This signaling, which is an attenuation of the earlier panic experience, is known as anxiety. In Freud's words, "Anxiety . . . is the expectation of the trauma on the one hand, and on the other, an attenuated repetition of it" (14, p. 114).

[6] The question of the first experience of anxiety and its effects on later experiences of anxiety has been one of controversy to the point of forming separate "schools" of psychoanalysis: the intense pain of hunger in the first few weeks of life, the shock of birth, or prenatal experiences have all been posited as the primary experience. All writers agree, however, that no matter what the initial circumstances might have been for an individual, later experiences of anxiety are in some sense referred to earlier ones.

The individual's problem, then, is to use the anxiety signal to become active and thus prevent the attenuated panic from developing into a full-blown one.

Since impulses continue to arise and tend toward motor expression, there is an ever recurrent pattern in the life of the individual: anxiety develops and attempts are instituted to prevent anxiety from growing into panic. The theory seems to imply (7, p. 143 f.) that under such circumstances new experiences do not function so as to modify infantile patterns of behavior and perception, but rather become assimilated into the memory traces connected with the impulse.[7] We must assume that any expression or recognition of the nature of the impulse or of the memory trace would lead to a recurrence of the original panic. We shall have more to say about this, but let us turn first to the role of the current situation.

The Role of the Current Situation [8]

At any given time, the momentary strength of an impulse is a function of two factors: (*a*) from within, it is related to impetus, which is a function of the state of deprivation of the organism, as described above; (*b*) from without, it is related to the "provocativeness" of situations for the expression of the impulse. The provocativeness of various situations for impulse expression is a difficult psychological problem. As we have pointed out in the above discussion of object, there may be cultural variations in the range of appropriate situations for impulse expression. Beyond these variations, there are modifications as a function of individual training. Nonetheless, it is necessary to assume that biologically some impulse-situation configurations are more "natural" than others, and that cultural modifications must confine themselves within certain limits. In the ecology of events in the environment, more and less provocative situations will occur for any given impulse. The greater the provocativeness of the situation, at any given degree of impetus, the greater will be the momentary impulse strength. As momentary impulse strength increases, the greater will be the tendency toward direct impulse expression, and the greater the energy required to inhibit direct expression.

Impulse expression does not occur in vacuo, but rather requires commerce with the environment. Thus people will seek out situations for impulse satisfaction when impetus increases. Where "natural" situations are unavailable, or where they are too anxiety provoking because of memory traces related to the impulse, then substitutions are found for the more natural ones.

Defense

There is a host of consequences to the inhibition of impulse expression. First, the impetus of the impulse will increase. With this increase in impetus there is an increasing likelihood of rearousal of panic if the latter is latent in the memory trace of the impulse. Second, the increased impetus demands increased effort on the part of the individual to prevent panic; in order to maintain equilibrium within the system,

[7] The theory is not clear on whether this course of events depends on the quantitative level of momentary anxiety and/or the intensity of the original panic, or whether it depends on the qualitative presence or absence of anxiety. The question is a crucial one for "ego psychology" (17).

[8] The current situation is rarely considered as a systematic variable in psychoanalytic theory. Its role is, however, clearly implied in many psychoanalytic writings. See, for example, the discussion on the role of the experiences of the previous day in the process of dream formation (9), and the effects of the current situation in relation to psychotherapy (16). The concept of momentary impulse strength does not appear in psychoanalytic writings, but seems useful to us.

energy must be expended for the prevention of panic. The energy bound up in this process, which has been called countercathexis, is not available for other organismic functions (7, p. 141 f.).

Let us consider the case where there is increased impetus and a "provocative" environmental situation. Let us consider, furthermore, that the experience of the individual (in the form of memory traces in the impulse presentation) has been such that greater increase in the momentary impulse strength will arouse panic. Anxiety signals this possibility of panic, and becomes a cue for the organism to institute defensive measures against panic. Such defensiveness may take many forms.[9] In some of these forms anxiety is allayed by avoiding recognition of or by distortion of the impulse; other forms function so as to avoid recognition of the provocativeness of the situation or by distortion of the situation. All defensive maneuvers to some extent and at some level avoid or distort the relationship between impulse and situation. When, as a function of anxiety, such distortions are necessary, new situations do not serve to alter perceptions or behavior in relation to impulses and their satisfaction.[10]

Behavioral Resultants: Derivatives

Having come by the devious route through discussion of impulse, memory trace, situation, anxiety, and defense, we are finally in a position to consider the behavior that results from the interplay of all these factors: the derivative of an unconscious conflict. In psychoanalysis behavior is seen either as rational, reality-oriented, appropriate, and goal-directed, or as irrational, irreality-oriented, inappropriate, and directed toward anxiety avoidance. Our judgments on these issues leave something to be wanted. Nevertheless, we can make approximations. Where behavior is consciously directed toward a goal; where it is integrated with other behaviors; where it is capable of modification with the acquisition of information; where interruption of the behavior leads to a search for other means-end objects or rational subgoals or substitute goals; and where completion of a behavior sequence results either in satisfaction or a modification of behavior, then we deem behavior to be appropriate. On the other hand, where an individual's behavior appears to be unintegrated and inconsistent with his other behaviors; where it is incapable of modification in the light of situational variants; where interruption leads to anxiety; and where completion of a behavior sequence leads neither to satisfaction nor to a modification of the sequence, we are led to believe such behavior to be a derivative of an unconscious conflict.[11]

Let us turn once more to our clinical example for concrete illustration of the abstractions we have been discussing. We do not know the specific natures of the impulses we are dealing with. We have evidence in the case of the student that there is a common impulse presentation that arises when he ap-

[9] The classical classification of defenses may be found in Anna Freud's *The Ego and the Mechanisms of Defence* (8).

[10] We have left the concepts of guilt and shame out of our discussion. These are often considered to be forms (with certain unique characteristics) that anxiety may take (7). These also lead to inhibitions and defenses.

[11] These listings are not to be construed as complete descriptions of "rational" and "irrational" behavior. Psychoanalytic theory has for the most part been concerned with irrational behavior. Freud himself was much interested in understanding rational behavior (cf. 20), and there are attempts currently at a psychoanalytic psychology of rational behavior (5; 17; 18; 22; 23, pp. 689–731). Recent psychoanalytic writings (cf. 17) imply that the distinction between rational and irrational processes and behavior is not as clear cut as we have suggested here.

pears to be seeking, or when situations call for responses of love, affection, sex, warmth in relationships. Such affects in the patient's past have been associated with panic. When in his early life wishes relating to these affects appeared, the patient's mother reacted with anxiety and treated him in such a way that psychological and physiological equilibrium could not be re-established. When he is currently tempted (either by recurrent impulse increases or by provocative situations) toward expressions in these areas, the memory trace of panic gives rise to anxiety. This the patient tries to allay by projection—he perceives others as making demands on him, and then complies with these perceived demands, but always in such a way as to create the impression that he is not doing his best. He thus leads people to believe that there are vast areas of untapped potentiality in him (as indeed there are), and they seek to guide and encourage him. In this manner he obtains some measure of love, while at the same time he is able to deny that he himself is seeking it. The behavior, as we see, exhibits both impulsive aspects (for example, seeking love) and defensive aspects (for example, projection).

By its very nature as the resultant of an unconscious impulse-defense conflict, a derivative does not provide true impulse satisfaction. At best, it lowers tension level momentarily.[12] Because of its unreality aspects, the production of derivatives tends to become progressively complicated, as for example when our patient feels resentment toward the people upon whom he has projected his own demands. The resentment against external demands has in its turn memory traces associated with it, in terms of the threat of loss of love. Anxiety arises when the expression of resentment is imminent, and further defensive measures are instituted by denial of affect. The resulting derivative behavior is often a superficial politeness which carries a tinge of supercilious condescension. The progressive development of this process has been referred to as the layering of defenses (**6, 7, 24**). Thus, the less satisfying the derivative: the greater the likelihood for more complicated defensive structure; the greater the amount of energy expended in maintaining the structure itself; the more numerous the areas of living permeated by pathology; and the less energy left over for other satisfactions.

[12] How a response which does not lead to direct physiological satisfaction of an impulse can lower tension level at all is unclear in the theory.

THE METHODOLOGY OF PSYCHOANALYTIC RESEARCH

Psychoanalysis developed rather independently of the tradition of experimental psychology. Psychology, in its academic setting and in its consciousness of itself as a new science, has consistently had the inclination and has found the time to examine the status of its methods. Psychoanalysis, on the other hand, developed in a clinical setting. The training of its workers, the daily exigencies of its practice, and the impelling force of the fruitfulness of its discoveries, have militated against a careful perusal of the assumptions and methods of its research. Though there are some exceptions, it is only recently that psychoanalytic writers have begun to turn from clinical problems back to the more general theoretical issues which so concerned Freud and his early coworkers (cf. **20**). If we give any credence to the discoveries of psychoanalysis, it is to the interest of not only clinical workers, but to all those concerned with the behavioral sciences, to look seriously at its methodology. For it would seem specious to accept its discoveries, even if only as a fruitful

source of hypotheses, and simultaneously dismiss the methods of making these discoveries. In this section we can only touch on a few of the methodological problems that grow out of our earlier discussion, and on some relationships between psychoanalytic and experimental research.

The Research Method

In common with other psychological systems, the observables for psychoanalytic theory are stimulus situations or the descriptions of stimulus situations and behavior or behavior descriptions. All other concepts are either hypothetical constructs or intervening variables. In order to point out the direction which psychoanalytic research takes within this framework, let us consider the nature of the therapist's activity as experimenter and empirical observer.

The patient describes, in the course of therapy, a wide variety of situations including the therapeutic one. Similarly, he describes and manifests a wide variety of behaviors. One assumption that the therapist works with in forming his hypotheses is that all situations that involve the same behavior are functionally equivalent. For example, if a patient responds to a variety of self-described situations or to the therapeutic situations with anger (or it may be twitching or weeping or withdrawal or descriptions of these), the therapist then forms a hypothesis which would relate these situations. By further observations, questions, and interpretations, he proceeds to check, modify, and extend his hypothesis. Thus he may observe that the behavior (or again, the patient's description of his behavior) of anger follows situations (or descriptions of situations) where affectionate feelings are aroused in the patient. He hypothesizes this relationship to be invariant. If the patient then expresses anger toward him, the therapist checks with the patient as to whether an unverbalized feeling of affection toward the therapist did not precede the angry outburst. Where his hypothesis fails to be confirmed, it is either completely or in some respect in error. Thus the therapist may, for example, modify the hypothesis to the form that there is an invariant relationship for this patient between affection-producing situations and anger in relationships with women but not with men, or through further checking, the modification may be of invariance in the case of women who behave in a seductive manner, but not toward those who behave in a nonseductive manner and not toward men. On the other hand, a more extensive revision of the hypothesis may be necessary. Such is the case, for example, should the patient become angry not only in one of the above described situations, but also, let us say, when the therapist fails to keep an appointment or is inattentive. We are now in the position of maintaining that there is a realm of situations (S_1, S_2 . . .) which invariantly leads to response R_1, and that there is at least one other realm of situations (S_A, S_B . . . , and perhaps S_α, S_β . . .) which also leads to response R_1, but that the former realm of situations is not equivalent to the latter realm. If we are to maintain our working assumption of the functional equivalence of all situations that involve the same behavior, then we must modify the hypothesis on the behavior side. Thus we may say that one realm of situations will produce response R_1 in a context of responses R_2, R_3 . . . , and that another realm of situations will produce response R_1 in a context of responses R_A, R_B . . . . Our definition of behavior now considers a series of responses.[13]

[13] The above discussion has dealt with the assumption of the psychological equivalence of objective situations leading to the same be-

The procedure we have discussed may at first glance seem to be overwhelmingly complicated. More than that, it may seem to the reader to be lacking in any sort of scientific rigor. Even were we to assume the objectivity of the therapist as observer, the pitfalls of the process are obvious. On the one hand our hypothesis can become so modified in terms of an increasingly refined description, that it serves to describe only a single event. As such, the hypothesis may be true but scientifically useless. On the other hand, it may become so broad in its extensions that it says nothing more than that people will demonstrate behavior in response to environmental events. As such, this may be true, but again scientifically useless. The problem in using such a procedure is one of achieving parsimony and specificity of prediction and is common to all science.

Unlike many of our more nomothetic investigations, the procedure is based on strict scientific determinism. It is similar, perhaps, to the chemical analysis of an unknown compound. Various hypotheses are brought forward, various reagents and tests are employed, hypotheses are modified and refined, and the goal is an accounting for all of the data with a minimal number of concepts. The table of elements, so to speak, and an accounting of their interactions are indeed unprecise in psychoanalytic theory, but there are guideposts for the observer. These guideposts are to be found in the theory discussed earlier in this paper, and in the genetic and structural theories of psychoanalysis. Thus the psychoanalytic observer can gauge to some extent via a succession of temporal events: the degree and type of defensiveness involved in a given derivative behavior, the similarity among behaviors, the points at which anxiety and defenses against anxiety appear, the kinds of impulses which may be broadly involved. From the patient's responses he can form tentative maps of the traces involved. From the genetic theory of such traces he derives hypotheses, and the empirical testing of these hypotheses leads to revision and refinement of the tentative mappings of traces. The investigation then extends to other patients and to the formulation of broader hypotheses. Ideally, at least, there is a constant interaction between theory and data.[14]

havior. Similarly, a series of hypotheses can in much the same way be based on the assumption that different behaviors following the same objective situation are functionally equivalent. Since we are speaking of temporal sequences, the word "same" with respect to behavior or situation always involves a proximate judgment of "roughly the same." We shall touch on this issue below. Both assumptions are utilized in the temporal process of psychotherapy.

The Individual Case as a Source of Data

As the above discussion implies, psychoanalysis is based upon research that involves the long-term study of individual cases. At first blush it would seem that the case study could not be considered research data in the usual sense, for even in these days of small sample techniques, our methods are not capable of handling samples as small as one. If we look at the individual case from the standpoint of derivative theory, however, the perspective changes markedly. Here we are interested in the relationships among impulse, memory trace, situation, anxiety, and defense, and within any one case repeated and stable patterns of these factors may be found.

Problems of the psychoanalytic case study. General issues regarding case study methods have been discussed in the literature (1, 21). We wish, how-

[14] A further discussion of the historical process of this interaction may be found in (19).

ever, to consider briefly some problems pertaining to the psychoanalytic case study as a method of developing and testing theories of behavior. There is one criticism frequently leveled against the use of such data for research purposes. First, since he has devoted so much time and energy to his training, the analyst looks for evidence to corroborate his favorite theory or subtheory, and he will naturally be able to find it in the mass of data which the patient produces. Worse still, since his beliefs are so strong, he will influence the patient, either consciously or unconsciously, to produce corroborative evidence. These are points on which Freud admonished his followers (and also himself) over 40 years ago (10, pp. 326–328), and they are still worth thinking about.

Freud's answer was that if the analyst is able to listen *freely* to the material presented by the patient (and he defined this sort of listening as a task corresponding to the patient's free association—and as full of as many pitfalls), then he will obtain free and uncontaminated data upon which to build and test theories. Freud specifically warns the analyst regarding research biases which may affect the ability to listen freely. The therapeutic analysis of the analyst is an attempt to guarantee free listening for both therapeutic and research purposes, but no one would claim that this guarantee has always worked out perfectly.

A further criticism is that the discovery of lawful regularities in the behavior of one individual is no indication that the same regularities will be found in another individual. Only further study can decide to what extent the functions described are universally relevant. This limitation, however, in no way affects the "scientific" legitimacy of the single case study for the formulation of psychological laws. We shall have more to say of this below.

The Units of Behavior

We have been speaking of associations, dreams, parapraxes, symptoms, and the like as the raw data of psychoanalytic research—in short, the behaviors that make up the "material" with which the psychoanalyst works. We shall now examine more closely the nature of these behaviors as research data.

Superficial consideration might lead one to think that the raw materials of psychoanalytic research are simple behaviors. What makes things more complicated, however, is that psychoanalysts are not interested in behaviors per se, but rather in what these behaviors *mean* for the patient. From the psychoanalytic viewpoint any isolated behavior is likely to be meaningless because it does not contain enough information to work with. The method that has been evolved is that of grouping behaviors into sequences and using the sequences themselves as basic data. Let us take a clinical example. A patient has headaches during her therapeutic sessions. The headaches are so severe that she cannot pay attention to anything else—she forgets what she has been discussing before, and she can hardly hear what the therapist is saying. Confronted solely with this behavior, neither the therapist nor the researcher can have much to say. If, however, one looks at the sequence of situations and behaviors in which headaches appear as a single behavioral event, the specific behavior takes on meaning. In one interview the patient talks about a man in whom she is interested, *then* gets a headache, *then* becomes angry that she cannot have a good time with people; in another interview she has physical sensations of warmth, feels light as though a weight

had been lifted from her shoulders, *then* gets a headache, *then* berates herself for not being able to talk with the therapist about problems which are frightening to her. In these and other instances of the sequence, the mood changes from mild euphoria and excitement in the first phase, to pain in the second, to depression in the third. As such sequences cumulate and are consistent in their course, and as further material is added, we gain increasing evidence that in this case the headache serves at least two functions, one to distract the patient from thinking about things which may provoke anxiety, and another (and at a deeper level) to punish her for thinking about forbidden things.

As we gather data on different patients who suffer from headaches, we may develop any of a number of hypotheses. One of these may be, for example, that headaches are a reflection of aggression turned inward. In testing such hypotheses we are no longer directly engaged in studying the processes involved in the occurrence of a specific behavior. We have in effect simply translated from behavior to impulse.[15] The question of the adequacy of this or any other symbolic translation is an empirical one. Its confirmation depends upon, among other things, an adequate sampling of a given class of people (in this case, those who have headaches). Tests of such hypotheses, while they are important to the study of the etiology of symptoms, tackle other problems in addition to those posed by the theory of derivatives. The relationship between behavior sequence and unconscious conflict, as exemplified by the above patient with headaches, may hold for one person only or for all people with a given symptom. Within the general theory, it is conceivable, for example, that headaches bear a relationship with aggression turned inward for one patient, a relationship with libidinization of visual activities for another patient, and an organic factor in a third. The problem is a legitimate empirical one, but only when intermediate steps between impulse and derivative are delineated will its investigation be crucial for psychoanalytic theory. Thus, for example, the fact that temperature elevation can be "caused" by any of a number of disease organisms, or can occur even in the absence of such organisms, does not "disprove" germ theory.[16]

Problems of behavior sequences. We turn now to some of the difficulties involved in studying sequences of behavior rather than isolated units of behavior. First, the very complexity of such sequences makes for research difficulties. While observational techniques may be developed with respectable degrees of reliability for identifying a single behavior, any unreliability that remains may be compounded if more than one behavior is studied at the same time. The reliability of judgment of the sequence itself, then, cannot be very high in the present state of the field. This difficulty is made even greater because many of the important sequences involve interaction between two or more people. To the extent that it provides

[15] The symbolic translation and the relationship to data are even more obscure when we select a group of subjects, all of whom have headaches, for studying the relationship between aggression turned inward and a third factor.

[16] A more directly psychoanalytic example is Erikson's (4) study of the Yurok, a tribe in which there is very compulsive concern with holding onto money. Rather than translating from concern with money to concern with feces, Erikson found the mode of "holding onto" associated with oral rather than anal training in Yurok childhood. Erikson is cognizant of the fact that in using a different cultural sample he has not refuted findings in Western European cultures, but rather he has extended the applicability of psychoanalytic theory.

multiple samples of sequences, the long period of time involved in the psychoanalytic treatment situation does help in the matter of practical reliability. But even here, as Freud (16) pointed out, certain situations and behaviors may not occur with sufficient frequency to be understood or dealt with therapeutically, within the course of a given analysis. If the researcher finds that the study of sequences of behavior answers important questions (or poses important problems), then he will devote more effort to improving techniques for studying them.

Another difficulty in the use of behavior sequences for research is that of selecting those sequences which one considers important. Even though psychoanalysts have been making use of patternings of behavior in their daily work for years, research is not oriented toward this sort of data, but rather toward the grouping of people according to classification systems that the researchers themselves claim to be outmoded. New classification schemes based on behavior sequences may eventually be developed if this approach seems fruitful, but in the meantime the selection of problems for research will probably be haphazard and based on personal predilections, just as the selection by the psychotherapist of behavior patterns for interpretation is to some extent based on personal theoretical likes and dislikes.

Finally, how much should be included in a behavior sequence for research? Should it include two contiguous behaviors, three, or how many? And how closely should recurrences of the sequence correspond before they are called identical? We have touched upon these matters in the previous section on research method, but they remain basic problems for psychology.

The Research Method as an Experimental Model

The research model utilizing successive hypotheses has rarely entered into psychological experiments. An experiment within this model would proceed through successive refinements of hypotheses and through successive tests of these hypotheses. *Any* discrepancies from predicted results would lead to a modification of relationships postulated among the variables studied, or to an additional concept. Such an experiment might be conducted with only two subjects, where the statistical test might be the probability of a given ordering among the subjects over a series of experimental situations. Any failure of the ordering would, however, lead to a re-examination of the hypotheses and to the introduction of a new series of experimental situations. From each succeeding series of experiments, the investigator would gain a more precise map of the operant variables. Such a procedure would in some ways resemble a chemical analysis, except that here we are interested in the relationships among variables rather than in the specific subjects, and furthermore, that limitations in our techniques require most often the ordinal handling of data. Aside from its demands on ingenuity in the selection and development of experimental tests as a function of preceding data, such a procedure is crude, and we do not know precisely where it would lead. We are examining our variables and our subjects simultaneously, but if, as we believe, this is true of all psychological investigations, then it is perhaps better that it be done systematically.

SUMMARY

In this paper we have tried to present an analysis of the structural relationships among concepts of psychoanalysis, as posed by the theory of derivatives. We have discussed aspects of the rela-

tionships among impulse, memory trace, anxiety, situation, and defense, and resultant derivative behavior. The final section of this paper considers the methodology from which this theory develops, and some implications of the theory and the methodology for research.

REFERENCES

1. ALLPORT, G. W. The use of personal documents in psychological science. *Soc. Sci. Res. Coun. Bull.*, 1942, No. 49.
2. BIBRING, E. The development and problems of the theory of the instincts. *Int. J. Psychoanal.*, 1941, 22, 102–132.
3. BLUM, G. S. *Psychoanalytic theories of personality*. New York: McGraw-Hill, 1953.
4. ERIKSON, E. H. Observations on the Yurok: childhood and world image. *Univer. Calif. Publ. Amer. Archaeol. Ethnol.*, 1943, **35**, 257–302.
5. ERIKSON, E. H. *Childhood and society*. New York: Norton, 1950.
6. FENICHEL, O. *Problems of psychoanalytic technique*. New York: The Psychoanalytic Quarterly, 1941.
7. FENICHEL, O. *The psychoanalytic theory of neurosis*. New York: Norton, 1945.
8. FREUD, ANNA. *The ego and the mechanisms of defence*. New York: International Universities Press, 1946.
9. FREUD, S. *The interpretation of dreams* (1900). Vol. V. *The complete psychological works of Sigmund Freud*. London: Hogarth, 1953.
10. FREUD, S. Recommendations for physicians on the psychoanalytic method of treatment (1912). In *Collected papers*. Vol. II. London: Hogarth, 1949. Pp. 323–334.
11. FREUD, S. Instincts and their vicissitudes (1915). In *Collected papers*. Vol. IV. London: Hogarth, 1949. Pp. 60–84.
12. FREUD, S. Repression (1915). In *Collected papers*. Vol. IV. London: Hogarth, 1949. Pp. 84–98.
13. FREUD, S. Two encyclopaedia articles (1922). In *Collected papers*. Vol. V. London: Hogarth, 1950. Pp. 107–136.
14. FREUD, S. *The problem of anxiety* (1926). New York: Norton, 1936.
15. FREUD, S. *New introductory lectures on psycho-analysis*. New York: Norton, 1933.
16. FREUD, S. Analysis terminable and interminable (1937). In *Collected papers*. Vol. V. London: Hogarth, 1950. Pp. 316–358.
17. HARTMANN, H. Ego psychology and the problem of adaptation (1939). In D. Rapaport (Ed.), *Organization and pathology of thought*. New York: Columbia Univer. Press, 1951. Pp. 362–396.
18. HARTMANN, H. Comments on the psychoanalytic theory of the ego. In *The psychoanalytic study of the child*. Vol. V. New York: International Universities Press, 1950. Pp. 74–97.
19. HARTMANN, H., KRIS, E., & LOEWENSTEIN, R. M. The function of theory in psychoanalysis. In R. M. Loewenstein (Ed.), *Drives, affects, behavior*. New York: International Universities Press, 1953. Pp. 13–38.
20. JONES, E. *The life and work of Sigmund Freud*. Vol. I. New York: Basic Books, 1953.
21. KELLY, E. L., & FISKE, D. W. *The prediction of performance in clinical psychology*. Ann Arbor: Univer. of Michigan Press, 1951.
22. KRIS, E. On preconscious mental processes. *Psychoanal. Quart.*, 1950, **19**, 540–560.
23. RAPAPORT, D. *Organization and pathology of thought*. New York: Columbia Univer. Press, 1951.
24. REICH, W. *Character-analysis*. New York: Orgone Institute, 1945.
25. STERBA, R. Introduction to the psychoanalytic theory of the libido. *Nerv. ment. Dis. Monogr.*, 1947, No. 68.

(Received January 27, 1954)

Psychological Review
Vol. 61, No. 6, 1954

A DECISION-MAKING THEORY OF VISUAL DETECTION [1]

WILSON P. TANNER, JR. AND JOHN A. SWETS
University of Michigan

This paper is concerned with the human observer's behavior in detecting light signals in a uniform light background. Detection of these signals depends on information transmitted to cortical centers by way of the visual pathways. An analysis is made of the form of this information, and the types of decisions which can be based on information of this form. Based on this analysis, the expected form of data collected in "yes-no" and "forced-choice" psychophysical experiments is defined, and experiments demonstrating the internal consistency of the theory are presented.

As the theory at first glance appears to be inconsistent with the large quantity of existing data on this subject, it is wise to review the form of these data. The general procedure is to hold signal size, duration, and certain other physical parameters constant, and to observe the way in which the frequency of detection varies as a function of intensity of the light signal. The way in which data of this form are handled implies certain underlying theoretical viewpoints.

In Fig. 1 the dotted lines represent the form of the results of hypothetical experiments. Consider first a single dotted line. Any point on the line might represent an experimentally determined point. This point is corrected for chance by application of the usual formula:

$$p = \frac{p' - c}{1 - c}, \qquad [1]$$

where p' is the observed proportion of positive responses, p is the corrected proportion of positive responses, and c is the intercept of the dotted curve at $\Delta I = 0$.

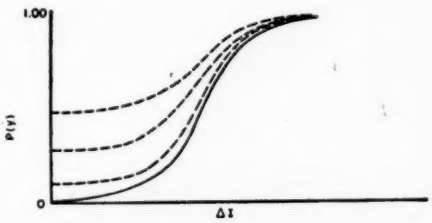

FIG. 1. Conventional seeing frequency or betting curve

Justification of this correction depends on the validity of the assumption that a "false alarm" is a guess, independent of any sensory activity upon which a decision might be based. For this to be the case it is necessary to have a mechanism which triggers when seeing occurs and which becomes incapable of discriminating between quantities of neural activity when seeing does not occur. Only under such a system would a guess be equally likely in the absence of seeing for all values of signal intensity. The application of the chance correction to data from both yes-no and forced-choice experiments is consistent with these assumptions.

The solid curve represents a "true" curve onto which each of the dotted, or experimental, curves can be mapped by using the chance correction and proper estimates of "c." The parameters of the solid curve are assumed to be characteristic of the physiology of the individual's sensory system, independent of psychological control. The assumption carries with it the

[1] This paper is based on work done for the U. S. Army Signal Corps under Contract No. DA-36-039 sc-15358. The experiments reported herein were conducted in the Vision Research Laboratories of the University of Michigan.

notion that if some threshold of neural activity is exceeded, phenomenal seeing results.

To infer that the form of the curve representing the frequency of seeing as a function of light intensity is the same as the curve representing the frequency of seeing as a function of neural activity is to assume a linear relationship between neural activity and light intensity. Efforts to fit seeing frequency curves by normal probability functions suggest a predisposition toward accepting this assumption.

A NEW THEORY OF VISUAL DETECTION

The theory presented in this paper differs from conventional thinking about these assumptions. First, it is assumed that false-alarm rate and correct detection vary together. Secondly, neural activity is assumed to be a monotonically increasing function of light intensity, not necessarily linear. A more specific statement than this is left for experimental determination.

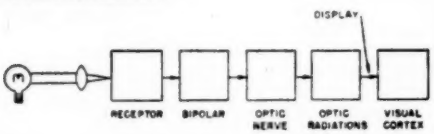

FIG. 2. Block diagram of the visual channel

Figure 2 is a block diagram of the visual pathways showing the major stages of transmission of visual information. All the stages prior to that labelled "cortex" are assumed to function only in the transmission of information, presenting to the cortex a representation of the environment. The function of interpreting this information is left to mechanisms at the cortical level.

In this simplified presentation, the displayed information consists of neural impulse activity. In the case under consideration, in which a signal is presented at a specified time in a known spatial location, the same restrictions are assumed to exist for the display. Thus, if the observer is asked to state whether a signal exists in location A at time B, he is assumed to consider only that information in the neural display which refers to location A at time B.

A judgment on the existence of a signal is presumably based on a

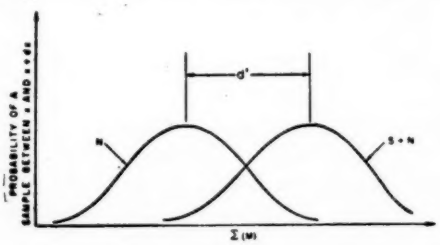

FIG. 3. Hypothetical distributions of noise and signal plus noise

measure of neural activity. There exists a statistical relationship between the measure and signal intensity. That is, the more intense the signal, the greater is the average of the measures resulting. Thus, for any signal there is a universe distribution which is in fact a sampling distribution. It includes all measures which might result if the signal were repeated and measured an infinite number of times. The mean of this universe distribution is associated with the intensity level of the signal. The variance may be associated with other parameters of the signal such as duration or size, but this is beyond the scope of this paper.

Figure 3 shows two probability distributions: N represents the case where noise alone is sampled—that is, no signal exists—and $S + N$, the case where signal plus noise exists. The mean of N depends on background

intensity; the mean of $S + N$ on background plus signal intensity. The variance of N depends on signal parameters, not background parameters in the case considered here; that is, where the observer knows a priori that if a signal exists then it is a particular signal. From the way the diagram is conceptualized, the greater the measure, $\Sigma(M)$, the more likely it is that this sample represents a signal. But one can never be sure. Thus, if an observer is asked if a signal exists, he is assumed to base his judgment on the quantity of neural activity. He makes an observation, and then attempts to decide whether this observation is more representative of N or of $S + N$. His task is, in fact, the task of testing a statistical hypothesis.

The ideal behavior, that which makes optimum use of the information available in this task, is defined mathematically by Peterson and Birdsall (2). The mathematics and symbols used are theirs, unless otherwise stated. The first case considered is the yes-no psychophysical experiment in which a signal is presented at a known location during a well-defined interval in time. This corresponds to Peterson and Birdsall's case of the signal known exactly.

For mathematical convenience, it is assumed that the distributions shown in Fig. 3 are Gaussian, with variance equal for N and all values of $S + N$. Experimental results suggest that equal variance is not a true assumption, but that the deviations are not great enough to justify the inconvenience of a more precise assumption for the purpose of this analysis.

It is also assumed that there is a cutoff point such that any measure of neural activity which exceeds that cutoff is in the criterion; that is, any value exceeding cutoff is accepted as representing the existence of a signal,

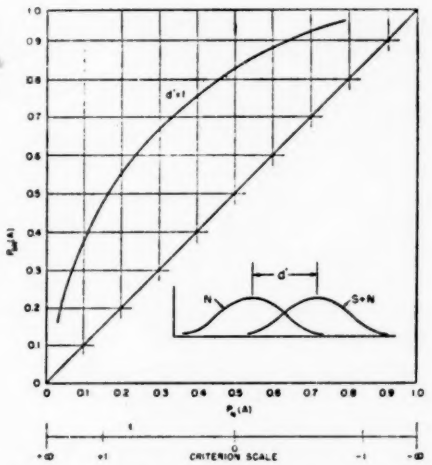

FIG. 4. $P_{SN}(A)$ vs. $P_N(A)$. The criterion scale shows the corresponding criteria expressed in terms of σ_N from M_N.

and any value less than the cutoff represents noise alone. Again, for mathematical convenience, the cutoff point is assumed to be well defined and stable. The justification for accepting this convenience is twofold: first, such behavior is statistically optimum, and second, if absolute stability is physically impossible, any lack of definition or random instability throughout an experiment has the same effect mathematically as additional variance in the sampling distributions.

Now, consider the way in which the placing of the cutoff affects behavior in the case of a given signal. In the lower right-hand corner of Fig. 4 the distributions N and $S + N$ are reproduced for a value of $d' = 1$. The parameter d' is the square root of Peterson and Birdsall's d. The square root of d is more convenient here; d' is the difference between the means of N and $S + N$ in terms of the standard deviation of N. The criterion scale is also calibrated in terms of the standard deviation of N. On the abscissa there is $P_N(A)$, the probability that, if no signal exists, the

measure will be in the criterion, and on the ordinate, $P_{SN}(A)$, the probability that if a signal exists, the measure will be in the criterion.

If the cutoff is at $-\infty$, all measures are in the criterion: $P_N(A) = P_{SN}(A) = 1$. At -1 standard deviation, $P_N(A) = .84$, and $P_{SN}(A) = .98$. At 0, $P_N(A) = .5$ and $P_{SN}(A) = .84$. At $+1$, $P_N(A) = .16$ and $P_{SN}(A) = .5$; and for $+\infty$ $P_N(A) = P_{SN}(A) = 0$. Thus, for $d' = 1$ this is the curve showing possible detections for each false-alarm rate. The curve represents the best that can be done with the information available, and the mirror image is the curve of worst possible behaviors.

The maximum behavior in any given experiment is a point on this curve at which the slope is β where

$$\beta = \frac{1-P(SN)}{P(SN)} \frac{(V_{N\cdot CA}+K_{N\cdot A})}{(V_{SN\cdot A}+K_{SN\cdot CA})}. \quad [2]$$

$P(SN)$ is the a priori probability that the signal exists, $V_{N\cdot CA}$ is the value of a correct rejection, $K_{N\cdot A}$ the cost of a false alarm, $V_{SN\cdot A}$ the value of a correct detection, and $K_{SN\cdot CA}$ is the cost of a miss. Thus, as $P(SN)$ or

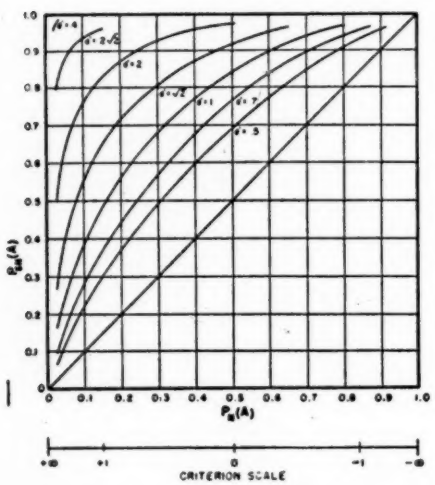

FIG. 5. $P_{SN}(A)$ vs. $P_N(A)$

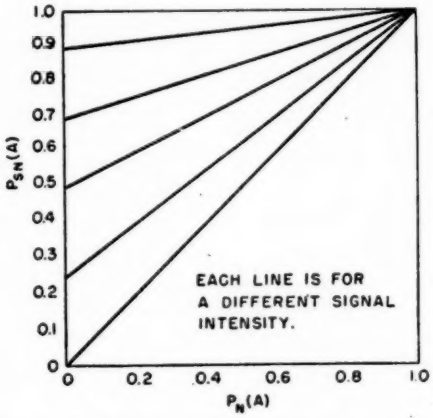

FIG. 6. $P_{SN}(A)$ vs. $P_N(A)$ as a function of d' assuming the guessing hypothesis

$V_{SN\cdot A}$ increases, or $K_{N\cdot A}$ decreases, β becomes smaller, and it is worth while to accept a higher false-alarm rate in the interest of achieving a greater percentage of correct decisions.

Figure 5 shows a family of curves of $P_{SN}(A)$ vs. $P_N(A)$ with d' as a parameter. For values of d' greater than 4, detection is very good. This is to be compared with the predictions of the conventional theory shown in Fig. 6 with $P_N(A)$ assumed to represent guesses. For each value of d' it is assumed that there is a true value of $P_{SN}(A)$ either for $P_N(A) = 0$ or for some very small value. The chance correction should transform each of these to horizontal lines.

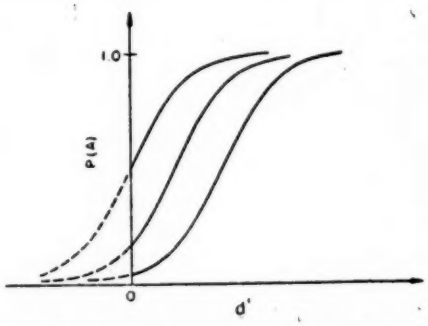

FIG. 7. $P(A)$ as a function of d' assuming the theory

Another way of comparing the predictions of this theory with those of conventional theory is to construct the so-called betting curves, or curves showing the predicted shape of the psychophysical function. These are shown in Fig. 7, where $P(A)$, the probability of acceptance, is plotted as a function of d'. These curves will not map onto the same curve by the application of the chance correction. The shift is horizontal rather than vertical. The dotted portions of the curve show that we are dealing with only a part of the curve, and thus, in the terms of this theory, it is improper to apply a normalizing procedure such as the chance-correction formula to that part of the curve.

In the forced-choice psychophysical experiment, maximum behavior is defined in a different way. In the general forced-choice experiment, the observer knows that the signal will occur in one of n intervals, and he is forced to choose in which of these intervals it occurs. The information upon which his decision is based is contained in the same display as in the case of the yes-no experiment, and, presumably, the values of d' for any given light intensity must be the same. While the solution of this problem is not contained in their study, Peterson and Birdsall have assisted greatly in determining this solution. The probability that a correct answer $P(C)$ will result for a given value of d' is the probability that one sample from the $S + N$ distribution is greater than the greatest of $n - 1$ samples from the distribution of noise alone. The case in which four intervals are used is the basis for Fig. 8. This figure shows the probability of one sample from $S + N$ being greater than the greatest of three from N. For a given value of d' this is

$$P(C) = \int_{x=-\infty}^{+\infty} F(x)^3 g(x) dx, \quad [3]$$

where $F(x)$ is the area of N and $g(x)$ is the ordinate of $S + N$. In Fig. 8 $P(C)$, as determined by this integration, is plotted as a function of d' for the equal-variance case.

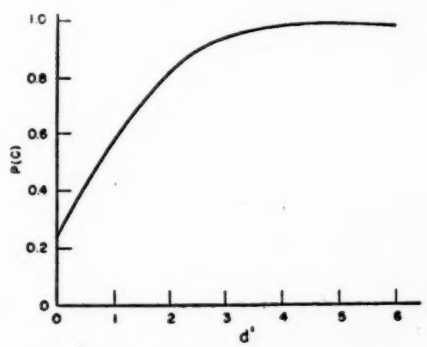

FIG. 8. $P(C)$ as a function of d'. A theoretical curve.

CRITERION OF INTERNAL CONSISTENCY

These two sets of predictions are for the standard experimental situations. They are based on the same neurological parameters. Thus, if the parameters, that is, d's, are estimated from one of the experiments, these estimates should furnish a basis for predicting the data for the other experiment if the theory is internally consistent. An equivalent criterion of internal consistency is that both experiments yield the same estimates of d'.

EXPERIMENTAL DESIGN

Experiments were conducted to test this internal consistency, using three Michigan sophomores as observers. All the experiments employed a circular target 30 minutes in diameter, 1/100 second in duration, on a 10-foot-lambert background. Details of the experimental procedure and the laboratory have been published by Blackwell, Pritchard, and Ohmart (1).

The observers were trained in the temporal forced-choice experiment. The signal appeared in a known location at one of four

specified times, and the observers were forced to choose the time at which they thought the signal occurred. Five light intensity increments were used here, with 50 observations per point per experimental session. The last two of these sessions were the test sessions, so that each forced-choice point in the analysis is based on 100 experimental observations.

Following the forced-choice experiments, there was a series of yes-no experiments under the same experimental conditions, except that only four light intensity increments were used. These were the same as the four greatest intensities used in the forced-choice experiments, reduced by adding a .1 fixed filter. In the first four of these sessions, two values of a priori probabilities, $P(SN)$ equal to .8 and .4, were used. The observers were informed of the value of $P(SN)$ before each experimental session. No values or costs were incorporated in these four sessions, which were thus excluded from the analysis as practice sessions.

The test experiments consisted of 12 sessions in each of which all of the information necessary for the calculation of a β (the best best possible decision level) was furnished the observers. While they did not know the formal calculation of β, that they knew the direction of cutoff change indicated by a change in any of these factors was suggested by the fact that the obtained values of $P_N(A)$ varied approximately with changes in the information given them. The values and costs were made real to the observers, for they were actually paid in cash. It was possible for them to earn as much as two dollars extra in a single experimental session as a result of this payment.

The first four sessions each carried the same value of β as $P(SN) = .8$ and the same payment was maintained. A high value of $P_N(A)$, or false-alarm rate, resulted. In the next four sessions with $P(SN)$ held at .8, $K_{N\cdot A}$ and $V_{N\cdot CA}$ were gradually increased from session to session (not within sessions) until $P_N(A)$ dropped to a low value. Then $P(SN)$ was dropped to .4, and K_{NA} and $V_{N\cdot CA}$ were reduced so that for the thirteenth session $P_N(A)$ stayed low. The last three sessions successively involved increases in $V_{S\cdot NA}$ and $K_{SN\cdot CA}$, again forcing $P_N(A)$ toward a higher value.

RESULTS

Figures 9 and 10 show scatter diagrams of $P_{SN}(A)$ vs. $P_N(A)$ for a particular intensity of signal and for a single observer. These scatter diagrams can be used to estimate d'. In Fig. 9 the estimate of d' is .7. In Fig. 10, the estimate of d' is 1.3. Each d' estimated in this way is based on 560 observations. A procedure similar to this was used for the d's for each of four signals for each of the four observers.

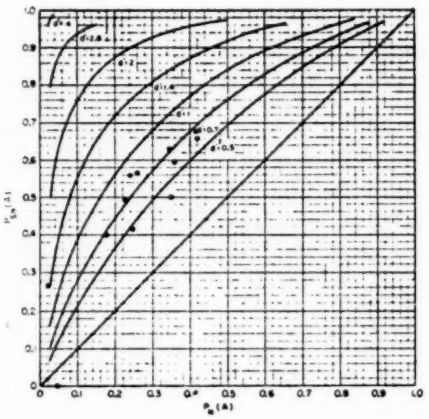

FIG. 9. A scatter diagram of $P_{SN}(A)$ vs. $P_N(A)$

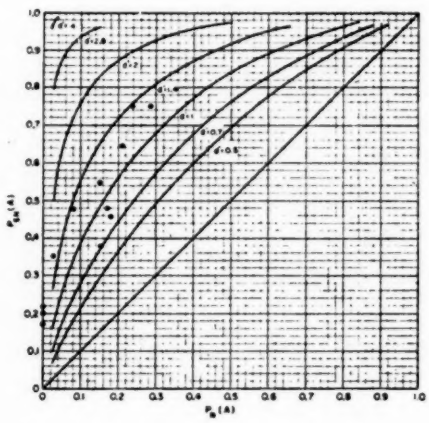

FIG. 10. A scatter diagram of $P_{SN}(A)$ vs. $P_N(A)$

In the forced-choice experiment the estimates of d' are made by entering our forced-choice curve (Fig. 8), using the observed percentage correct as an estimate of $P(C)$. Figure 11 shows log d' as a function of log signal in-

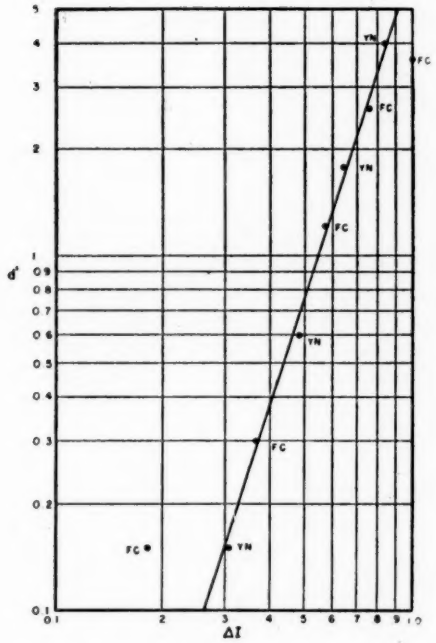

FIG. 11. Log d' vs. Log ΔI for Observer 1

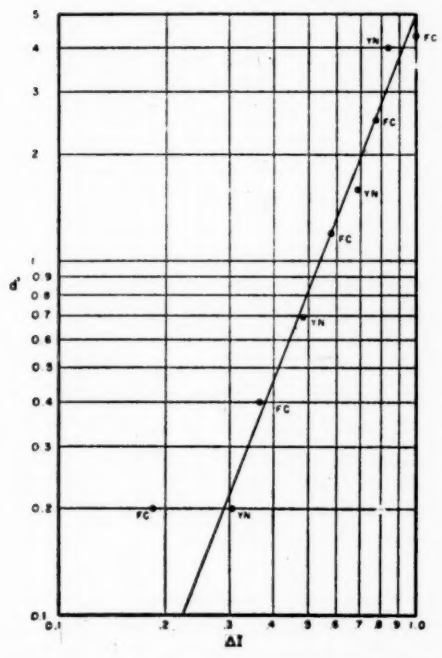

FIG. 12. Log d' vs. Log ΔI for Observer 2

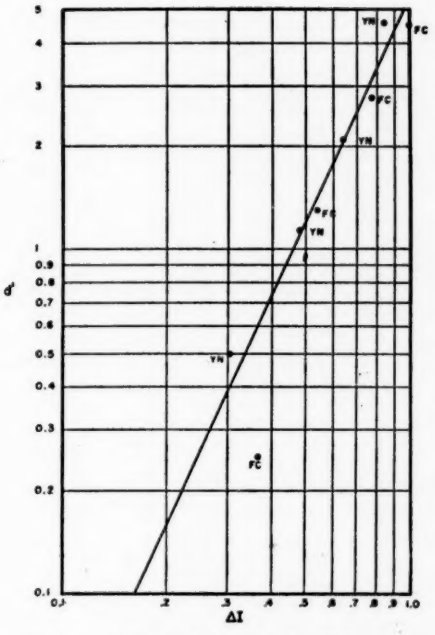

FIG. 13. Log d' vs. Log ΔI for Observer 3

tensity for the first observer, the estimates of d' being from both forced-choice and yes-no experiments. In general the agreement is good. The deviation of the forced-choice point at the top can be explained on the basis of inadequate experimental data for the determination of the high probability involved. The deviation of the low point is unexplained. Figure 12 is the same plot for the second observer, showing about the same picture. Figure 13 is for the third observer, showing not quite as good a fit, but nevertheless satisfactory for psychological experiments. For this observer, the lowest point for forced choice is off the graph to the right of the line.

Figures 14, 15, and 16 show the predictions for forced-choice data (when yes-no data are used to estimate d') for the three observers. Note that the lowest point is on the curve in both of the first two cases,

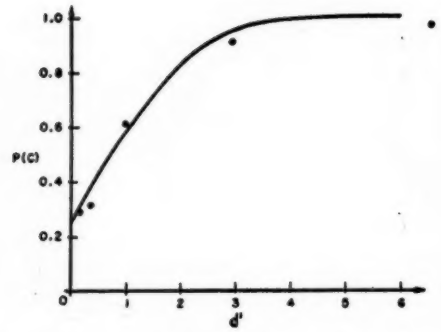

FIG. 14. Prediction of forced-choice data from yes-no data for Observer 1

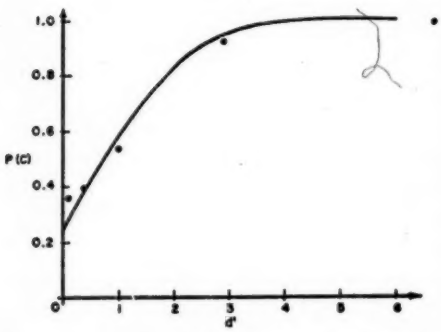

FIG. 15. Prediction of forced-choice data from yes-no data for Observer 2

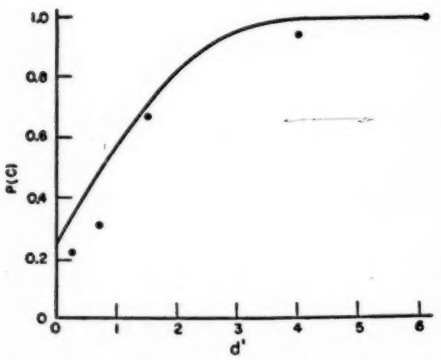

FIG. 16. Prediction of forced-choice data from yes-no data for Observer 3

suggesting that the deviation which appeared on the curves in Fig. 11, 12, and 13 is not significant.

DISCUSSION

The results satisfy the criterion of internal consistency. The theory also turns out to be consistent with the vast amount of data in the literature, for, when the d' vs. ΔI function for any one of the observers is used to predict probability of detection as a function of ΔI in terms of this theory, the result closely approximates the type of curve frequently reported. Shapes of curves thus furnish no basis for selecting between the two theories, and a decision must rest on the other arguments.

According to conventional theory, application of the chance correction should yield corrected values of $P_{SN}(A)$ which are independent of $P_N(A)$, or should yield corrected thresholds in the conventional sense which are independent of $P_N(A)$. Rank-order correlations for the three observers between $P_N(A)$ and corrected thresholds (.30, .71, .67) are highly significant; the combined $p \ll .001$. This is a result consistent with theory presented here.

Another method of comparison is to fit the scatter diagrams (Fig. 9 and 10) by straight lines. According to the independence theory, these straight lines should intercept the point (1.00, 1.00). Sampling error would be expected to send some of the lines to either side of this point. There are 12 of these scatter diagrams, and all 12 of these lines intersect the line $P_{SN}(A) = 1.00$ at values of $P_N(A)$ between 0 and 1.00 in an order which would be predicted if these lines were arcs of the curves $P_{SN}(A)$ vs. $P_N(A)$ as defined by the theory of signal detectability.

Two additional sessions were run in which the observers were permitted three categories of response (yes, no, and doubtful), and were told to be sure of being correct if they responded

either yes or no. Again, two a priori probabilities (.8 and .4) were employed, and again $P_N(A)$ was correlated with $P(SN)$. The observers, interviewed after these sessions, reported that their "yes" responses were based on "phenomenal" seeing.

This does not mean that the observers were abnormal because they hallucinated. It suggests, on the other hand, that phenomenal seeing develops through experience, and is subject to change with experience. Psychological as well as physiological factors are involved. Psychological "set" is a function of β, and after experience with a given set one begins to see, or not to see, rather automatically. Change the set, and the level of seeing changes. The experiments reported here were such that the observers learned to adjust rapidly to different sets.

CONCLUSIONS

The following conclusions are advanced: (*a*) The conventional concept of a threshold, or a threshold region, needs re-evaluation in the light of the present theory that the visual detection problem is the problem of detecting signals in noise. (*b*) The hypothesis that false alarms are guesses is rejected on the basis of statistical tests. (*c*) Change in neural activity is a power function of change in light intensity. (*d*) The mathematical model of signal detection is applicable to problems of visual detection. (*e*) The criterion of seeing depends on psychological as well as physiological factors. In the experiments reported here the observers tended to use optimum criteria. (*f*) The experimental data support the assumption of a logical connection between forced-choice and yes-no techniques developed by the theory.

REFERENCES

1. BLACKWELL, H. R., PRITCHARD, B. S., & OHMART, T. G. Automatic apparatus for stimulus presentation and recording in visual threshold experiments. *J. opt. Soc. Amer.*, 1954, **44**, 322–326.
2. PETERSON, W. W., & BIRDSALL, T. G. The theory of signal detectability. Electronic Defense Group, Univer. of Michigan, *Tech. Rep.*, No. 13, Sept., 1953.

(Received October 29, 1953)

Psychological Review
Vol. 61, No. 6, 1954

ON ABELSON'S CRITICAL COMMENT

DAVID BAKAN

University of Missouri

Abelson's paper (1) leads me to conclude that either the original paper (2) is very unclear, or Abelson misunderstood it, or both. Under any circumstances, given what seems to be his understanding of it, he has indeed been very gracious in his criticism; for that which he attributes to the paper makes it patently absurd.

Essentially he has ascribed his own definitions, and at least one additional assumption, to the original development and, very understandably, finds that this leads to difficulties.

In the original paper $P(g)$ is used in the sense of the *strength* of g, in much the same way as Thorndike thought of the strength of a connection (2, p. 368). If g is the "what is learned" (2, p. 362), $P(g)$ is the degree of it. It appears likely, particularly in view of Abelson's misinterpretation, that this was perhaps not sufficiently explicated in the original paper.

Abelson defines $P(g)$ as "the probability that the organism is in the condition g" (1, p. 276). It is difficult to conceive of a definition more at variance with both the intent and the content of the original paper. By his definition the locus of $P(g)$ is *outside* the organism. He is distinctly in error when he attributes an outside-the-organism $P(g)$, in this simple sense, to "Bakan's model" (1, p. 278). The outside-the-organism $P(g)$ stems rather from *his* definition of $P(g)$.

A similar straw man which he sets up for criticism is contained in his phrase, "If we take seriously Bakan's formulation, which allows only two conditions for the organism, g and 'not-g' . . ." (1, p. 277). It is certainly true that the paper centers around g and $\bar{g}$, but no one has ever concluded that there are only two kinds of animals because the animal universe may be divided into horses and not-horses. No two-g assumption is made in the paper at all. The conclusions that Abelson derives based on this assumption are interesting, but they are his own.

REFERENCES

1. ABELSON, R. B. Critical comment on "Learning and the principle of inverse probability." *Psychol. Rev.*, 1954, **61**, 276–278.
2. BAKAN, D. Learning and the principle of inverse probability. *Psychol. Rev.*, 1953, **60**, 360–370.

(Received July 6, 1954)

PSYCHOLOGICAL REVIEW

July 1, 1954

YEAR	VOLUME	AVAILABLE NUMBERS	PRICE PER NUMBER	PRICE PER VOLUME
1894	1	– 2 – 4 5 6	$1.50	——
1895	2	– – 3 4 5 6	$1.50	——
1896	3	– – – – – –	——	——
1897	4	1 – – – – 6	$1.50	——
1898	5	– – 3 – 5 –	$1.50	——
1899	6	– – – – – –	$1.50	——
1900	7	1 – – – – –	$1.50	——
1901	8	1 2 – – – –	$1.50	——
1902	9	– 2 – – – –	$1.50	——
1903	10	1 2 – – – –	$1.50	——
1904	11	1 – – 4 & 5 6	$1.50	——
1905	12	1 2 & 3 4 5 –	$1.50	——
1906	13	– – 3 4 5 6	$1.50	——
1907	14	1 2 – – – –	$1.50	——
1908	15	– – – – – –	——	——
1909	16	1 – 3 4 5 –	$1.50	——
1910	17	1 2 3 – – 6	$1.50	——
1911	18	1 2 3 4 5 –	$1.50	——
1912	19	1 2 3 4 5 6	$1.50	$8.00
1913	20	1 2 3 4 5 6	$1.50	$8.00
1914	21	– 2 3 4 5 6	$1.50	——
1915	22	1 2 3 4 5 6	$1.50	$8.00
1916	23	1 – – 4 5 –	$1.50	——
1917	24	– – – 4 – –	$1.50	——
1918	25	– 2 3 4 5 6	$1.50	——
1919	26	1 2 3 4 5 6	$1.50	$8.00
1920	27	1 – – – – –	$1.50	——
1921	28	– 2 – – – 6	$1.50	——
1922	29	1 – – 4 – –	$1.50	——
1923	30	1 – 3 – – –	$1.50	——
1924	31	1 2 3 4 5 6	$1.50	$8.00
1925	32	– 2 3 – 5 –	$1.50	——
1926	33	1 – 3 4 5 6	$1.50	——
1927	34	1 2 3 4 5 6	$1.50	$8.00
1928	35	1 2 3 4 5 6	$1.50	$8.00
1929	36	1 2 3 4 5 6	$1.50	$8.00
1930	37	1 2 3 4 5 6	$1.50	$8.00
1931	38	1 2 3 4 5 6	$1.50	$8.00
1932	39	1 2 3 4 5 6	$1.50	$8.00
1933	40	1 2 3 4 5 6	$1.50	$8.00
1934	41	1 2 3 4 5 6	$1.50	$8.00
1935	42	1 2 3 4 5 6	$1.50	$8.00
1936	43	1 2 3 4 5 6	$1.50	$8.00
1937	44	1 2 3 4 5 6	$1.50	$8.00
1938	45	1 2 3 4 5 6	$1.50	$8.00
1939	46	1 2 3 4 5 6	$1.50	$8.00
1940	47	– – 3 4 – –	$1.50	——
1941	48	– – – – – 6	$1.50	——
1942	49	– 2 3 4 5 6	$1.50	——
1943	50	1 2 3 4 5 6	$1.50	$8.00
1944	51	1 2 3 4 5 6	$1.50	$8.00
1945	52	1 2 3 4 – 6	$1.50	——
1946	53	1 2 3 4 5 6	$1.50	$8.00
1947	54	1 2 – 4 5 6	$1.50	——
1948	55	1 2 3 4 5 6	$1.50	$8.00
1949	56	1 2 3 4 5 6	$1.50	$8.00
1950	57	1 2 3 4 5 6	$1.50	$8.00
1951	58	1 2 3 4 5 6	$1.50	$8.00
1952	59	1 2 3 4 5 6	$1.50	$8.00
1953	60	1 2 3 4 5 6	$1.50	$8.00
1954	61	By subscription $6.50, foreign $7.00	$1.50	——

Postage prepaid on U. S. orders. Add $.50 per volume on foreign orders. All stock subject to prior sale.
The American Psychological Association gives the following discounts on orders for any one journal:

10% on orders of $ 50.00 and over
20% on orders of $100.00 and over
30% on orders of $150.00 and over

Current subscriptions and orders for back numbers should be addressed to

AMERICAN PSYCHOLOGICAL ASSOCIATION
1333 Sixteenth Street N.W. Washington 6, D. C.

PSYCHOLOGICAL MONOGRAPH

#270

1945

ON PROBLEM SOLVING

By

KARL DUNCKER

$2.50

This popular monograph is #270 of the Psychological Monograph series. It has been reprinted so that it is again available.

Third Printing

AMERICAN PSYCHOLOGICAL ASSOCIATION

1333 Sixteenth Street N. W.

Washington 6, D. C.